

RAN 4227/58

5

Biotin Biosynthetic GenesBackground of the Invention

10 The present invention relates to the production process of biotin by fermentation using a genetically engineered organism.

Biotin is one of the essential vitamins for nutrition of animals, plants, and microorganisms, and very important as medicine or food additives.

15

Biotin biosynthesis of *Escherichia coli* has been studied well, and it has been clarified that biotin is synthesized from pimelyl CoA via 7 keto-8-amino pelargonic acid (KAPA), 7,8-diamino pelargonic acid (DAPA) and desthiobiotin (DTB) [*Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology, 544, (1987)]. The analysis of genetic information involved in the biosynthesis of biotin has been advanced on *Escherichia coli* [J. Biol. Chem., 263, 19577, (1988)] and *Bacillus sphaericus* (US Patent No. 5096823). At least four enzymes are known to be involved in this biosynthetic pathway. These four enzymes are encoded by the bioA, bioB, bioD and bioF genes. The bioF gene codes for KAPA synthetase which catalyzes the conversion of pimelyl CoA to KAPA. The bioA gene codes for DAPA aminotransferase which converts KAPA to DAPA. The bioD gene codes for DTB synthetase which converts DAPA to DTB. The bioB gene codes for biotin synthase which converts DTB to biotin. It has been also reported that the bioC and bioH genes are involved in the synthesis of pimelyl CoA in *Escherichia coli*.

20

25

There are many studies on fermentative production of biotin. *Escherichia coli* (Japanese Patent Kokai No. 149091/1986 and Japanese Patent Kokai No. 155081/1987), *Bacillus sphaericus* (Japanese Patent Kokai No. 180174/1991), *Serratia marcescens* (Japanese Patent Kokai No. 27980/1990) and *Brevibacterium flavum* (Japanese Patent Kokai No. 240489/1991) have been used. But these processes have not yet been suitable for use in an industrial production process because of a low productivity. Moreover, large amounts of DTB, a biotin precursor, accumulates in the fermentation of these bacteria. Therefore, it has been assumed that the last step of the biotin biosynthetic pathway, from DTB to biotin, is a rate limiting step.

30

35

On the other hand, it was found that a bacterial strain belonging to the genus *Kurthia* produces DTB and small amounts of biotin. Also mutants which produce much larger amounts of biotin were derived from wild type strains of the genus *Kurthia* by selection for resistance to biotin antimetabolites acidomycin (ACM), 5-(2-thienyl)-valeric acid (TVA) and alpha-methyl desthiobiotin (MeDTB). However, in view of the still low biotin titers it is desirable to apply genetic engineering to improve the biotin productivity of such mutants.

### Summary of the Invention

The present invention relates therefore to the chromosomal DNA fragments carrying the genes involved in the biotin biosynthesis of *Kurthia* sp.. The isolated chromosomal DNA fragments carry 8 genes, the *bioA*, *bioB*, *bioC*, *bioD*, *bioF*, *bioFII*, *bioH* and *bioHII* genes, and transcriptional regulatory sequences. The *bioFII* gene codes for an isozyme of the *bioF* gene product. The *bioHII* gene codes for an isozyme of the *bioH* gene product.

The present invention further relates to *Kurthia* sp. strains in which at least one gene involved in biotin biosynthesis is amplified, and also to the production process of biotin by this genetically engineered *Kurthia* sp. strain.

Although the DNA fragment mentioned above may be of various origins, it is preferable to use the strains belonging to the genus *Kurthia*. Specific examples of such strains include, for example, *Kurthia* sp. 538-6 (DSM No. 9454) and its mutant strains by selection for resistance to biotin antimetabolites such as *Kurthia* sp. 538-KA26 (DSM No. 10609).

### Brief Description of the Figures

Before the present invention is explained in more detail by referring to the following examples a short description of the enclosed Figures is given:

Fig. 1: Restriction maps of pKB100, pKB200 and pKB300.

Fig. 2: Structure of pKB100.

Fig. 3: Structure of pKB200.

Fig. 4: Restriction maps and complementation results of pKH100, pKH101 and pKH102.

Fig. 5: Structure of pKH100.

Fig. 6: Restriction maps of pKC100, pKC101 and pKC102.

Fig. 7: Structure of pKC100.

Fig. 8: Structures of derived plasmids from pKB100, pKB200 and pKB300.

Fig. 9: Gene organizations of the gene clusters involved in biotin biosynthesis of *Kurthia* sp. 538-KA26.

5 Fig. 10: Nucleotide sequence between the ORF1 and ORF2 genes of *Kurthia* sp. 538-KA26.

Fig. 11: Nucleotide sequence of the promoter region of the *bioH* gene cluster.

Fig. 12: Nucleotide sequence of the promoter region of the *bioFII* gene cluster.

Fig. 13: Construction of the shuttle vector pYK1.

10 Fig. 14: Construction of the *bioB* expression plasmid pYK114.

### Detailed Description of the Invention

15 Generally speaking the present invention is directed to DNA molecules comprising polynucleotides encoding polypeptides represented by SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14 or 16, and functional derivatives of these polypeptides which contain addition, insertion, deletion and/or substitution of one or more amino acid residue(s), and to DNA molecules comprising polynucleotides which hybridize under stringent hybridizing conditions to polynucleotides  
20 which encode such polypeptides and functional derivatives. The invention is also directed to vectors comprising one or more such DNA sequences, for example, a vector wherein said DNA sequences are functionally linked to promoter sequence(s). The invention is further directed to biotin-expressing cells, said cells having been transformed by one or more DNA sequences or vector(s) as defined above, and a process for the production of biotin which comprises  
25 cultivating a biotin-expressing cell as defined above in a culture medium to express biotin into the culture medium, and isolating the resulting biotin from the culture medium by methods known in the art. Any conventional culture medium and culturing conditions may be used in accordance with the invention. Preferably, such a process is carried out wherein the cultivation is effected from 1 to 10 days, preferably from 2 to 7 days, at a pH from 5 to 9, preferably from  
30 6 to 8, and a temperature range from 10 to 45°C, preferably from 25 to 30°C.

Finally, the present invention is also directed to a process for the preparation of pharmaceutical, food or feed compositions characterized therein that biotin obtained by such processes is mixed with one or more generally used additives with which a man skilled in the art  
35 is familiar.

The DNA molecules of the invention may be produced by any conventional means, such as by the techniques of genetic engineering and automated gene synthesis known in the art.

5 A detailed method for isolation of DNA fragments carrying the genes coding for the enzymes involved in the biotin biosynthesis from these bacterial strains is described below.

Therefore, DNA can be extracted from Kurthia sp. 538-KA26 by the known phenol method. Such DNA is then partially digested by Sau3AI and ligated with pBR322 digested by  
10 BamHI to construct a genomic library of Kurthia sp. 538-KA26.

Biotin auxotrophic mutants which lack the biosynthetic ability to produce biotin are transformed with the genomic library obtained above, and transformants showing biotin prototrophy are selected. The selected transformants have the genomic DNA fragments  
15 complementing deficient genes in the biotin auxotrophic mutants. As biotin auxotrophic mutants, Escherichia coli R875 (bioB<sup>-</sup>), R877 (bioD<sup>-</sup>), BM7086 (bioH<sup>-</sup>) and R878 (bioC<sup>-</sup>) (J. Bacteriol., 112, 830-839, (1972) and J. Bacteriol., 143, 789-800, (1980) can be used. The transformation of such Escherichia coli strains can be carried out according to a conventional method such as the competent cell method [*Molecular Cloning*, Cold Spring Harbor Laboratory  
20 Press, 252, (1982)].

In the present invention, a hybrid plasmid which complements the bioB deficient mutant of Escherichia coli was obtained in the manner described above. The obtained hybrid plasmid is named pKB100. The pKB100 corresponds to plasmid pBR322 carrying a 5.58 Kb of a  
25 genomic DNA fragment from Kurthia sp. 538-KA26, and its restriction cleavage map is shown in Fig. 1 and 2.

The hybrid plasmid named pKB200 which complements the bioD deficient mutant of Escherichia coli was also obtained as described above. The pKB200 corresponds to plasmid  
30 pBR322 carrying a 7.87 Kb of genomic DNA fragment from Kurthia sp. 538-KA26, and its restriction cleavage map is shown in Fig. 1 and 3. The genomic DNA fragment in pKB200 completely overlaps with the inserted fragment of the pKB100 and carries the bioF, bioB, bioD, ORF1 and ORF2 genes and a part of the bioA gene of Kurthia sp. 538-KA26 as shown in Fig. 9-A.

35

The complete bioA gene of Kurthia sp. 538-KA26 can be isolated by conventional methods, such as colony hybridization using a part of the genomic DNA fragment in pKB200 as a probe. The whole DNA of Kurthia sp. 538-KA26 is digested with a restriction enzyme such as HindIII and ligated with a plasmid vector cleaved by the same restriction enzyme. Then, Escherichia coli is transformed with the hybrid plasmids carrying genomic DNA fragments of Kurthia sp. 538-KA26 to construct a genomic library. As a vector and Escherichia coli strain, the pUC19 [Takara Shuzo Co.(Higashiiru, Higashinotohin, Shijodohri, Shimogyo-ku, Kyoto-shi, Japan)] and Escherichia coli JM109 (Takara Shuzo Co.) can be used, respectively.

The hybrid plasmid named pKB300 carrying a 8.44 Kb genomic DNA fragment from Kurthia sp. 538-KA26 was obtained by colony hybridization and its restriction cleavage map is shown in Fig. 1. The genomic DNA fragment in the pKB300 carries two gene clusters involved in the biotin biosynthesis of Kurthia sp. 538-KA26 as shown in Fig. 9-A. One cluster consists of the ORF1, bioD and bioA genes. Another cluster consists of the ORF2, bioF and bioB genes. The nucleotide sequences of the bioD and bioA genes are shown in SEQ ID No. 1 and SEQ ID NO: 3, respectively. The predicted amino acid sequences of the bioD and bioA gene products are shown in SEQ ID NO: 2 and SEQ ID NO: 4, respectively. The bioD gene codes for a polypeptide of 236 amino acid residues with a molecular weight of 26,642. The bioA gene codes for a polypeptide of 460 amino acid residues with a molecular weight of 51,731. The ORF1 gene codes for a polypeptide of 194 amino acid residues with a molecular weight of 21,516, but the biological function of this gene product is unknown.

The nucleotide sequences of the bioF and bioB genes are shown in SEQ ID NO: 5 and SEQ ID NO: 7, respectively. The predicted amino acid sequences of the bioF and bioB gene products are SEQ ID NO: 6 and SEQ ID NO: 8, respectively. The bioF gene codes for a polypeptide of 387 amino acid residues with a molecular weight of 42,619. The bioB gene codes for a polypeptide of 338 amino acid residues with a molecular weight of 37,438. The ORF2 gene codes for a polypeptide of 63 amino acid residues with a molecular weight of 7,447, but the biological function of this gene product is unknown. Inverted repeat sequences which are transcriptional terminator signals are found downstream of the bioA and bioB genes. As shown in Fig. 10, two transcriptional promoter sequences which initiate transcriptions in both directions are found between the ORF1 and ORF2 genes. Furthermore, there are two inverted repeat sequences named Box1 and Box2 involved in the negative control of the transcriptions between each promoter sequence and each translational start codon.

In addition, two hybrid plasmids which complement the biotin auxotrophic mutants of *Escherichia coli* were obtained in the manner described above. The hybrid plasmid named pKH100 complements the bioH deficient mutant, and the hybrid plasmid named pKC100 the bioC mutant. pKH100 (Fig. 4 and 5) has a 1.91 Kb genomic DNA fragment from *Kurthia* sp. 538-KA26 carrying a gene cluster consisting of the bioH and ORF3 genes as shown in Fig. 9-B. The nucleotide sequence of the bioH gene and the predicted amino acid sequence of this gene product are shown in SEQ ID NO: 9 and SEQ ID NO: 10, respectively. The bioH gene codes for a polypeptide of 267 amino acid residues with a molecular weight of 29,423. The ORF3 gene codes for a polypeptide of 86 amino acid residues with a molecular weight of 9,955, but the biological function of this gene product is unknown. A promoter sequence is found upstream of the bioH gene as shown in Fig. 11, and there is an inverted repeat sequence which is the transcriptional terminator downstream of the ORF3 genes. Since the promoter region has no inverted sequence such as Box1 and Box2, it is expected that the expressions of these genes are not regulated.

On the other hand, pKC100 carries a 6.76 Kb genomic DNA fragment from *Kurthia* sp. 538-KA26 as shown in Fig. 6 and 7. The genomic DNA fragment in pKC100 carries a gene cluster consisting of the bioFII, bioHII and bioC genes as shown in Fig. 9-C. The bioHII and bioFII genes are genes for isozymes of the bioH and bioF genes, respectively, because the bioHII and bioFII genes complement the bioH deficient and the bioF deficient mutants of *Escherichia coli*, respectively. The nucleotide sequences of the bioFII, bioHII and bioC genes are shown in SEQ ID NO: 11, SEQ ID NO: 13 and SEQ ID NO: 15, respectively. The predicted amino acid sequences of the bioFII, bioHII and bioC gene products are shown in SEQ ID NO: 12, SEQ ID NO: 14 and SEQ ID NO: 16, respectively. The bioFII gene codes for a polypeptide of 398 amino acid residues with a molecular weight of 44,776. The bioHII gene codes for a polypeptide of 248 amino acid residues with a molecular weight of 28,629. The bioC gene codes for a polypeptide of 276 amino acid residues with a molecular weight of 31,599. A promoter sequence is found upstream of the bioFII gene, and there is an inverted repeat sequence named Box3 in the promoter region as shown in Fig. 12. The transcription of these genes terminates at an inverted repeat sequence existing downstream of the bioC gene. Since the nucleotide sequence of Box3 is significantly similar to those of Box1 and Box2, expressions of these genes is estimated to be regulated similarly to the bioA and bioB gene clusters.

Needless to say, the nucleotide sequences and amino acid sequences of the genes isolated above are artificially changed in some cases, e.g., the initiation codon GTG or TTG may be converted into an ATG codon.

5        Therefore the present invention is also directed to functional derivatives of the polypeptides of the present case. Such functional derivatives are defined on the basis of the amino acid sequence of the present invention by addition, insertion, deletion and/or substitution of one or more amino acid residues of such sequences wherein such derivatives still have the same type of enzymatic activity as the corresponding polypeptides of the present invention. Such activities can  
10       be measured by any assays known in the art or specifically described herein. Such functional derivatives can be made either by chemical peptide synthesis known in the art or by recombinant means on the basis of the DNA sequences as disclosed herein by methods known in the state of the art, such as, e.g., that disclosed by Sambrook et al. (*Molecular Cloning*, Cold Spring Harbour Laboratory Press, New York, USA, second edition 1989). Amino acid exchanges in  
15       proteins and peptides which do not generally alter the activity of such molecules are known in the state of the art and are described, for example, by H. Neurath and R.L. Hill in "The Proteins" (Academix Press, New York, 1979, see especially Figure 6, page 14). The most commonly occurring exchanges are: Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Thy/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu,  
20       Asp/Gly as well as the reverse.

      Furthermore the present invention is not only directed to the DNA sequences as disclosed e.g., in the sequence listing as well as their complementary strands, but also to those which include these sequences, DNA sequences which hybridize under Standard Conditions with such  
25       sequences or fragments thereof and DNA sequences, which because of the degeneration of the genetic code, do not hybridize under Standard Conditions with such sequences but which code for polypeptides having exactly the same amino acid sequence.

      "Standard Conditions" for hybridization mean in this context the conditions which are  
30       generally used by a man skilled in the art to detect specific hybridization signals and which are described, e.g., by Sambrook et al., (s.a.) or preferably so-called stringent hybridization and non-stringent washing conditions, or more preferably so-called stringent hybridization and stringent washing conditions a man skilled in the art is familiar with and which are described, e.g., in Sambrook et al. (s.a.).

35

DNA sequences which are derived from the DNA sequences of the present invention either because they hybridize with such DNA sequences (see above) or can be constructed by the polymerase chain reaction by using primers designed on the basis of such DNA sequences can be prepared either as indicated namely by the PCR reaction, or by site directed mutagenesis [see  
5 e.g., Smith, Ann. Rev. Genet. 19, 423 (1985)] or synthetically as described, e.g., in EP 747 483 or by the usual methods of *Molecular Cloning* as described, e.g., in Sambrook et al. (s.a.).

As a host strain for the expression and/or amplification of the DNA sequences of the present invention, any microorganism may be used, e.g., those identified in EP 635 572, but it  
10 is preferable to use the strains belonging to the genus *Kurthia*, especially *Kurthia* sp. 538-6 (DSM No. 9454) and *Kurthia* sp. 538-51F9 (DSM No. 10610).

In order to obtain a transformant with a high biotin productivity, the DNA sequences of the present invention are used under the control of a promoter which is effective in such host cells.  
15 The DNA sequences of the present invention can be introduced into the host cell by transformation with a plasmid carrying such DNA sequences or by integration into the chromosome of the host cell.

When *Kurthia* sp. 538-51F9 is used as the host cell, *Kurthia* sp. 538-51F9 may be  
20 transformed with a hybrid plasmid carrying at least one gene involved in biotin biosynthesis isolated above from a *Kurthia* sp. strain. As a vector plasmid for the hybrid plasmid, pUB110 [J. Bacteriol., 154, 1184-1194, (1983)], pHP13 (Mol. Gen. Genet., 209, 335-342, 1987), or other plasmids comprising the origin of replication functioning in *Kurthia* sp. strain can be used. As DNA sequences for amplification and/or expression in *Kurthia* sp. any DNA sequence of the  
25 present invention can be used, but the DNA sequence corresponding to the bioB gene coding for biotin synthase is preferred. One example of such a hybrid plasmid is pYK114 shown in Fig. 14. In this plasmid the bioB gene is under the control of the promoter for the bioH gene and carries the replicating origin of pUB110.

30 *Kurthia* sp. 538-51F9 may be transformed with pYK114 obtained as described above by the protoplast transformation method [Molecular Biological Methods for *Bacillus*, 150, (1990)]. However, since *Kurthia* sp. 538-51F9 has a low efficiency of regeneration from protoplasts, transformation efficiency of this strain is very low. Therefore, it is preferred to use a strain having high efficiency of regeneration from protoplasts should be used, e.g., *Kurthia* sp. 538-  
35 51F9-RG21 which can be prepared as described in Example 14 of the present case.



The present invention also provides a process for the production of biotin by the cultivation of the thus obtained transformants, and separation and purification of the produced biotin.

5 Cultivation of the biotin-expressing cells of the present invention can be done by methods known in the art. The culturing conditions are not critical so long as they are sufficient for the expression of biotin by the biotin-expressing cells to occur. A culture medium containing an assimilable carbon source, a digestible nitrogen source, an inorganic salt, and other nutrients necessary for the growth of the biotin-expressing cell can be used. As the carbon source, for  
10 example, glucose, fructose, lactose, galactose, sucrose, maltose, starch, dextrin or glycerol may be employed. As the nitrogen source, for example, peptone, soybean powder, corn steep liquor, meat extract, ammonium sulfate, ammonium nitrate, urea or a mixture thereof may be employed. Further, as an inorganic salt, sulfates, hydrochlorides or phosphates of calcium, magnesium, zinc, manganese, cobalt and iron can be employed. And, if necessary,  
15 conventional nutrient factors or an antifoaming agent, such as animal oil, vegetable oil or mineral oil can also be added. If the obtained biotin-expressing cell has an antibiotic resistant marker, the respective antibiotic should be supplemented into the medium. The pH of the culture medium may be between 5 to 9, preferably 6 to 8. The cultivation temperature can be 10 to 45°C, preferably 25 to 30°C. The cultivation time can be 1 to 10 days, preferably 2 to 7 days.

20

The biotin produced under the conditions as described above can easily be isolated from the culture medium by methods known in the art. Thus, for example, after solid materials have been removed from the culture medium by filtration, the biotin in the filtrate may be absorbed on active carbon, then eluted and purified further with an ion exchange resin. Alternatively, the  
25 filtrate may be applied directly to an ion exchange resin and, after the elution, the biotin is recrystallized from a mixture of alcohol and water.

## Examples

### Example 1

Cloning bioB and bioF genes of *Kurthia* sp. 538-KA26.

5

#### 1. Preparation of the genomic library.

The acidomycin-resistant strain of *Kurthia* sp., 538-KA26 (DSM No. 10609), was cultivated in 100 ml of nutrient broth (Kyokuto Seiyaku Co; Honcho 3-1-1, Nihonbashi, Chuoh-Ku, Tokyo, Japan) at 30°C overnight, and bacterial cells were recovered by  
10 centrifugation. The whole DNA was extracted from the bacterial cells by the phenol extraction method [Experiments with gene fusions, Cold Spring Harbor Laboratory, 137-138, (1984)], and 1.9 mg of the whole DNA was obtained.

The whole DNA (10 µg) was partially digested with 1.2 units of Sau3AI at 37°C for 1  
15 hour to yield fragments with around 10 Kb in length. 5-15 Kb DNA fragments were obtained by agarose gel electrophoresis.

The vector pBR322 (Takara Shuzo Co.) was completely digested with BamHI, and then treated with alkaline phosphatase to avoid self ligation. The DNA fragments were ligated with  
20 the cleaved pBR322 using a DNA ligation Kit (Takara Shuzo Co.) according to the instruction of the manufacturer. The ligation mixture was transferred to *Escherichia coli* JM109 (Takara Shuzo Co.) by the competent cell method [*Molecular Cloning*, Cold Spring Harbor Laboratory, 252-253, (1982)], and the strains were selected for ampicillin resistance (100 µg/ml) on agar plate LB medium (1% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl, pH 7.5). About  
25 5,000 individual clones having the genomic DNA fragments were obtained as a genomic library.

The ampicillin-resistant strains of the genomic library of *Kurthia* sp. 538-KA26 were cultivated at 37°C overnight in 50 ml of LB medium containing 100 µg/ml ampicillin, and bacterial cells were collected by centrifugation. Plasmid DNA was extracted from the bacterial  
30 cells by the alkaline-denaturation method [*Molecular Cloning*, Cold Spring Harbor Laboratory, 90-91, (1982)].

## 2. Selection of the clone carrying the bioB gene from the genomic library.

The plasmid DNA was transferred by the competent cell method into *Escherichia coli* bioB deficient mutant R875 (J. Bacteriol. 112, 830-839, 1972) without a biotin synthetase activity.

- 5 The transformed *Escherichia coli* R875 cells were washed twice with 0.85 % NaCl and streaked on 1.5% agar plates of M9CT medium (0.6%  $\text{Na}_2\text{HPO}_4$ , 0.3%  $\text{HK}_2\text{PO}_4$ , 0.05% NaCl, 0.1%  $\text{NH}_4\text{Cl}$ , 2 mM  $\text{MgSO}_4$ , 0.1 mM  $\text{CaCl}_2$ , 0.2% glucose, 0.6 % vitamin-free casamino acid, 1  $\mu\text{g/ml}$  thiamin) containing 100  $\mu\text{g/ml}$  of ampicillin, and the plates were incubated at 37°C for 40 hours. One transformant with the phenotype of the biotin prototrophy was obtained. The
- 10 transformant was cultivated in LB medium containing 100  $\mu\text{g/ml}$  ampicillin, and the hybrid plasmid was extracted from the cells. The hybrid plasmid carries an insert of 5.58 Kb and was designated pKB100. The restriction map is shown in Fig. 1 and 2.

## 3. Complementation of biotin deficient mutants of *Escherichia coli* with pKB100.

15

- pKB100 was transferred to biotin deficient mutants of *Escherichia coli*, R875 (bioB<sup>-</sup>), W602 (bioA<sup>-</sup>), R878 (bioC<sup>-</sup>), R877 (bioD<sup>-</sup>), R874 (bioF<sup>-</sup>) or BM7086 (bioH<sup>-</sup>) [J. Bacteriol., 112, 830-839, (1972) and J. Bacteriol., 143, 789-800, (1980)], by the competent cell method. The transformed mutants were washed with 0.85% NaCl three times and plated on M9CT agar
- 20 plates containing 100  $\mu\text{g/ml}$  of ampicillin and 0.1 ng/ml biotin, and the plates were incubated at 37°C overnight. Colonies on the plates were replicated on M9CT agar plates with 100  $\mu\text{g/ml}$  ampicillin in the presence or absence of 0.1 ng/ml biotin, the plates were incubated at 37°C for 24 hours to perform the complementation analysis. As shown in Table 1, the pKB100 could complement not only the bioB but also the bioF mutant. In contrast, bioA, bioC, bioD and bioH
- 25 mutants were not complemented by pKB100. From this results, it was confirmed that the pKB100 carried the bioB and bioF genes of *Kurthia* sp. 538-KA26.

Table 1

Plasmid	Escherichia coli biotin deficient mutant					
	bioA <sup>-</sup>	bioB <sup>-</sup>	bioC <sup>-</sup>	bioD <sup>-</sup>	bioF <sup>-</sup>	bioH <sup>-</sup>
pKB100	-	+	-	-	+	-
pKB200	-	+	-	+	-	-
pKB300	+					
pKH100	-	-	-	-	-	+
pKC100	-	-	+	-	+	+

Example 2

5

Isolation of hybrid plasmid carrying of the bioD gene of Kurthia sp. 538-KA26.

1. Isolation of the hybrid plasmid carrying the bioD gene.

10

The genomic library of Kurthia sp. 538-KA26 of Example 1-1 was transferred into the Escherichia coli bioD deficient mutant R877, and transformants having an ampicillin resistance and biotin prototrophy phenotype were selected in the same manner as described in Example 1-2. The transformant were cultivated at 37°C overnight in LB medium with 100 µg/ml ampicillin, and the bacterial cells were collected by centrifugation. The hybrid plasmid was extracted from the cells by the alkaline-denaturation method. The hybrid plasmid had a 7.87 Kb insert DNA fragment and was designated pKB200. Cleavage patterns of pKB200 were analyzed using various restriction endonucleases (HindIII, NcoI, EcoRI, BglII, SalI, and PstI) and compared with that of pKB100. Restriction endonuclease analysis revealed that the two hybrid plasmids had exactly the same cleavage sites and that the 1.5 Kb DNA fragment was extended to the left side of pKB100 and the 0.8 Kb fragment was stretched out to the right side in the pKB200 (Fig. 1 and 3).

15

20

2. Complementation of biotin deficient mutant of Escherichia coli with pKB200.

25

The pKB200 was transferred to the biotin deficient mutants of Escherichia coli, R875 (bioB<sup>-</sup>), W602 (bioA<sup>-</sup>), R878 (bioC<sup>-</sup>), R877 (bioD<sup>-</sup>), R874 (bioF<sup>-</sup>) or BM7086 (bioH<sup>-</sup>). Complementation analysis was performed by the method described in Example 1-3. The pKB200 complemented the bioD and bioB mutants, but not the bioA, bioC, bioF and bioH

mutants as shown in Table 1. Although the pKB200 overlapped on the whole length of pKB100, pKB200 did not complement the bioF mutant.

### Example 3

Isolation of the hybrid plasmid carrying the bioH gene of Kurthia sp. 538-KA26.

#### 1. Isolation of the hybrid plasmid carrying the bioH gene.

5  
10 The genomic library of Kurthia sp. 538-KA26 of Example 1-1 was transferred to the Escherichia coli bioH deficient mutant BM7086. Transformants having the bioH clone were selected for biotin prototrophy in the same manner as in Example 1-2. The hybrid plasmid were extracted from the transformed cells by the alkaline-denaturation method and analyzed by restriction enzymes. The hybrid plasmid had 1.91 Kb inserted DNA fragment and was  
15 designated pKH100. Since the genomic library used above has 5-15 Kb of the genomic DNA fragments, the pKH100 was thought to be subjected to a modification, such as deletion, in Escherichia coli strain. The restriction map of the pKH100 is shown in Fig. 4 and 5. Cleavage patterns of the pKH100 were completely different from those of pKB100 and pKB200. Therefore, pKH100 carried a DNA fragment of the Kurthia chromosome which differed from  
20 those in pKB100 and pKB200.

#### 2. Complementation of the biotin deficient mutant of Escherichia coli with pKH100.

Complementation analysis was performed by the method described in Example 1-3. The  
25 pKH100 was transferred to the biotin deficient mutants of Escherichia coli, R875 (bioB<sup>-</sup>), W602 (bioA<sup>-</sup>), R878 (bioC<sup>-</sup>), R877 (bioD<sup>-</sup>), R874 (bioF<sup>-</sup>) or BM7086 (bioH<sup>-</sup>). pKH100 complemented only the bioH mutant, but not the bioB, bioA, bioC, bioD and bioF mutants as shown in Table 1. Thus, pKH100 carries the bioH gene.

#### Example 4

Isolation of the hybrid plasmid carrying the bioC gene of Kurthia sp. 538-KA26.

##### 5 1. Isolation of the hybrid plasmid carrying the bioC gene.

The genomic library of Kurthia sp. 538-KA26 of Example 1-1 was transferred to the Escherichia coli bioC deficient mutant R878. Transformants with the bioC clone were selected for biotin prototrophy in the same manner as described in Example 1-2. The hybrid plasmid was  
10 extracted from the transformant cells by the alkaline-denaturation method and analyzed with restriction enzymes. The hybrid plasmid had a 6.76 Kb inserted DNA fragment and was designated pKC100. The restriction map of pKC100 is shown in Fig. 6 and 7. Cleavage patterns of pKC100 were completely different from those of pKB100, pKB200 and pKH100. Therefore, pKC100 carries a different region of the Kurthia chromosome from those of  
15 pKB100, pKB200 and pKH100.

##### 2. Complementation of the biotin deficient mutant of Escherichia coli with pKC100.

The complementation analysis was performed by the method described in Example 1-3. pKC100 was transferred to the biotin deficient mutants of Escherichia coli, R875 (bioB<sup>-</sup>), W602 (bioA<sup>-</sup>), R878 (bioC<sup>-</sup>), R877 (bioD<sup>-</sup>), R874 (bioF<sup>-</sup>) or BM7086 (bioH<sup>-</sup>). pKC100 complemented the bioC, bioF and bioH mutants as shown in Table 1. Since the inserted DNA fragment in pKH100 was different from those in pKB100 and pKH100, pKC100 carried not only the bioC gene but also genes for isozymes of the bioF gene product (KAPA synthetase)  
25 and the bioH gene product.

#### Example 5

Isolation of the hybrid plasmid carrying the bioA gene of Kurthia sp. 538-KA26.

30

##### 1. Isolation of the left region of the chromosomal DNA in pKB200.

We isolated the left region of the chromosomal DNA in pKB200 from Kurthia sp. 538-KA26 chromosomal DNA by the hybridization method. The whole DNA of Kurthia sp. 538-KA26 was completely digested with HindIII and subjected to agarose gel electrophoresis. The  
35

DNA fragments on the gel were denatured and then transferred to a nylon membrane (Hybond-N, Amersham) according to the recommendations of the manufacturer.

pKB200 was completely digested with NcoI, and a 2.1 Kb NcoI fragment was isolated by agarose gel electrophoresis (Fig. 1). The NcoI fragment was labeled with  $^{32}\text{P}$  by the Multiprime DNA labeling system (Amersham) and used as a hybridization probe. The hybridization was performed on the membrane prepared above using the "Rapid hybridization buffer" (Amersham) according to the instructions of the manufacturer. The probe strongly hybridized to a HindIII fragment of about 8.5 Kb.

In order to isolate the 8.5 Kb fragment, the whole DNA of Kurthia sp. 538-KA26 was completely digested with HindIII, and 7.5-9.5 Kb DNA fragments were obtained by agarose gel electrophoresis. The vector plasmid pUC19 (Takara Shuzo Co.) was completely digested with HindIII and treated with alkaline phosphatase to avoid self ligation. The 7.5-9.5 Kb DNA fragments were ligated with the cleaved the pUC19 using a DNA ligation Kit (Takara Shuzo Co.), and the reaction mixture was transferred to Escherichia coli JM109 by the competent cell method. About 1,000 individual clones carrying such genomic DNA fragments were obtained as a genomic library.

The selection was carried out by the colony hybridization method according to the protocol described by Maniatis et al. [*Molecular Cloning*, Cold Spring Harbor Laboratory, 312-328, (1982)]. The grown colonies on the agar plates were transferred to nylon membranes (Hybond-N, Amersham) and lysed by alkali. The denatured DNA was immobilized on the membranes.  $^{32}\text{P}$  labeled NcoI fragments prepared as described above were used as a hybridization probe, and the hybridization was performed using the "Rapid hybridization buffer" (Amersham) according to the instructions of the manufacturer. Three colonies which hybridized with the probe DNA were obtained, and hybrid plasmids in these colonies were extracted by the alkaline-denaturation method.

The structure analysis was performed with restriction enzymes (BamHI, HindIII, NcoI, EcoRI, BglII, SalI and PstI ). All of the three hybrid plasmids had a 8.44 Kb inserted DNA fragment, and the three hybrid plasmids had exactly the same cleavage patterns. These results indicated that they were identical. This hybrid plasmid was designated pKB300. The restriction map of pKB300 is shown in Fig. 1. About half length of the genomic DNA fragment in pKB300 overlaps with that of pKB200.

## 2. Complementation of the bioA deficient mutant of Escherichia coli with pKB300.

5 The complementation analysis of Escherichia coli W602 (bioA<sup>-</sup>) with pKB300 was performed by the method described in Example 1-3. Since pKB300 complemented the bioA mutation (Table 1), pKB300 carries the bioA gene of Kurthia sp..

### Example 6

10 Subcloning of the bioA, B, D and F genes of Kurthia sp. 538-KA26.

#### 1. Construction of the hybrid plasmid pKB103 and pKB104.

15 pKB100 was completely digested with HindIII, and a 3.3 Kb HindIII fragment was isolated. The 3.3 Kb fragment was ligated with the vector pUC18 (Takara Shuzo Co.) cleaved with HindIII using a DNA ligation Kit to construct the hybrid plasmids pKB103 and pKB104. In pKB103 and pKB104, the 3.3 Kb fragments were inserted in both orientations relative to the promoter-operator of the lac gene in pUC18. Their restriction map is shown in Fig. 8.

20 Complementation of the bioB or bioF deficient mutants of Escherichia coli (R875 or R874) were performed with pKB103 and pKB104 in the same manner as described in Example 1-3. pKB103 and pKB104 complemented the bioB and bioF mutants (Table 2).

#### 2. Construction of derivatives of pKB200.

25

Since pKB200 complemented the bioD mutation and covered the whole length of pKB100 carrying the bioB and bioF genes, a series of deletion mutations of pKB200 were constructed to localize more precisely bioB, bioD and bioF. A 4.0 Kb Sall-HindIII fragment of pKB200 was inserted into the Sall and HindIII sites of pUC18 and pUC19 to give pKB221 and pKB222 in  
30 which the Sall-HindIII fragment is placed in both orientations.

pKB200 was completely digested with NruI, and a 7.5 Kb NruI fragment was isolated by agarose gel electrophoresis. The NruI fragment was recirculated by the DNA ligation Kit, and pKB223 was obtained.

35



pKB200 was completely digested with HindIII. A 4.8 Kb HindIII fragment was isolated by agarose gel electrophoresis and cloned into the HindIII site of pUC18 in both orientations to generate pKB224 and pKB225.

5 pKB200 was partially digested with NcoI, and a 3.1 Kb NcoI fragment was isolated by agarose gel electrophoresis. The ends of the NcoI fragment were made blunt by using the Klenow fragment of the DNA polymerase I (Takara Shuzo Co.) and ligated with HindIII linker (Takara Shuzo Co.) The 3.1 Kb HindIII fragment was obtained by treatment with HindIII and cloned into the HindIII site of the pUC19 in both orientation to give pKB228 and pKB229.

10 In the same manner, both ends of a 2.1 Kb NcoI fragment of pKB200 were converted to HindIII sites by treatment with the Klenow fragment and addition of HindIII linkers. Then the obtained HindIII fragment was inserted into the HindIII site of pUC19 in both orientations to give pKB230 and pKB231.

15 pKB234 and pKB235 were generated by insertion of a 1.6 Kb HindIII-NruI fragment of pKB230 into the HindIII and SmaI sites of pUC19 and pUC18, respectively.

20 The restriction maps of the pKB200 derivatives are shown in Fig. 8.

### 3. Complementation analysis of biotin deficient mutants of Escherichia coli with pKB200 derivatives.

25 Complementation analysis was performed with the pKB200 derivatives in the same manner as described in Example 1-3. The complementation results are summarized in Table 2. The bioB deficient mutant was complemented by pKB221, pKB222, pKB224 and pKB225, but not by pKB223, pKB228, pKB229, pKB230, pKB231, pKB234 and pKB235. The bioF deficient mutant was complemented by pKB223, pKB224, pKB225, pKB228 and pKB229, 30 but not by pKB221, pKB222, pKB230, pKB231, pKB234 and pKB235. On the other hand, the bioD deficient mutant was complemented by pKB223, pKB224 and pKB225, but not by pKB221 and pKB222.

35 Together with the complementation analysis with pKB103 and pKB104, these results support that the bioF gene is present at the left side of the first NruI site on pKB103 while the

bioB gene is located on the right side of the same NruI site with a short overlap to the left and that the bioD gene is present on at most 1.5 Kb left side region of the pKB200. Thus, the complementation results with various derivatives of pKB100 and pKB200 showed that the bioD, bioF and bioB genes lie in turn on the 4.4 Kb region at the left side of the HindIII site of pKB200.

#### 4. Construction of the hybrid plasmid pKB361.

To determine the location of the bioA gene, the derivative of pKB300 was constructed. pKB361 was generated by insertion of a 2.8 Kb BamHI-SalI fragment of pKB300 into the BamHI and SalI sites of pUC19 (Fig. 8).

pKB361 was transferred to the bioA deficient mutant of Escherichia coli (W602), and complementation analysis was performed in the same manner as described in Example 1-3. The bioA mutant was complemented by pKB361 (Table 2), suggesting the presence of the bioA gene within the 2.8 Kb region between the BamHI and SalI sites of pKB300.

Table 2

Plasmid	Escherichia coli biotin deficient mutant			
	bioA <sup>-</sup>	bioD <sup>-</sup>	bioF <sup>-</sup>	bioB <sup>-</sup>
pKB103, 104			+	+
pKB221, 222		-	-	+
pKB223		+	+	-
pKB224, 225		+	+	+
pKB228, 229			+	-
pKB230, 231			-	-
pKB234, 235			-	-
pKB361	+			

### Example 7

Subcloning of the bioH gene of Kurthia sp. 538-KA26.

5      1. Construction of the hybrid plasmids pKH101 and pKH102.

pKH100 was completely digested with BamHI and recirculated with a DNA ligation Kit to generate pKH101 in which a 0.75 Kb BamHI fragment was deleted from pKH100 (Fig. 4).

10      pKH102 was constructed from pKH100 by treatment with HindIII followed by recirculation with a DNA ligation Kit. The pKH102 lacked a 1.07 Kb HindIII fragment in pKH100 (Fig. 4).

15      Complementation analysis of the Escherichia coli bioH mutant (R878) was performed with pKH101 and pKH102 in the same manner as in Example 1-3. pKH101 complemented the bioH mutant, but not pKH102 (Fig. 4). This result indicated that the bioH gene is located in the left region (1.16 Kb) of the BamHI site on pKH100.

### Example 8

Subcloning of the bioC gene of Kurthia sp. 538-KA26.

20

pKC100 was completely digested with BamHI, and a 1.81 Kb BamHI fragment was isolated by agarose gel electrophoresis. The BamHI fragment was ligated with pBR322 treated with BamHI and the Klenow fragment by a DNA ligation Kit. Finally, pKC101 and pKC102 in which the BamHI fragment was inserted in both orientations were obtained (Fig. 6).

25

pKC101 and pKC102 were transferred to the Escherichia coli bioC mutant R878, and complementation analysis was carried out in the same manner as in Example 1-3. The bioC mutant was complemented with pKC101 and pKC102, and the bioC gene was confirmed to lie in the 1.81 Kb BamHI fragment.

30

### Example 9

Nucleotide sequence of the inserted DNA fragments on pKB100, pKB200 and pKB300.

5 For nucleotide sequencing analysis of the inserted DNA fragments of pKB100, pKB200 and pKB300, several subclones overlapping mutually were constructed using pUC18, pUC19, M13mp18 and M13mp19 (Takara Shuzo Co.) and a series of deletion derivatives of the subclones were obtained by the Kilo-Sequencing Deletion Kit (Takara Shuzo Co.). Then, nucleotide sequencing analysis of the deletion derivatives was carried out by the dideoxy-chain  
10 termination technique (Sequenase version 2.0 DNA sequencing kit using 7-deaza-dGTP, United States Biochemical Co.). The results were analyzed by the computer program (GENETYX) from Software Development Co..

Computer analysis of this sequence revealed that the cloned DNA fragment has the capacity  
15 to code for six open reading frames (ORF). This gene operon has two gene clusters proceeding to both directions (Fig. 9-A).

The first ORF in the left gene cluster starts with the TTG codon preceded by a ribosomal binding site (RBS) with homology to the 3' end of the *Bacillus subtilis* 16S rRNA and codes for  
20 a protein of 194 amino acid residues having a molecular weight of 21,516. It was not possible to determine the function of the gene product by the complementation analysis, accordingly, this ORF was named ORF1.

The nucleotide sequence of the second ORF in the left gene cluster is shown in SEQ ID  
25 NO: 1. This gene codes for a protein of 236 amino acid residues with a molecular weight of 26,642. The predicted amino acid sequence of this gene product is shown in SEQ ID NO: 2. A putative RBS is found upstream of the ATG initiation codon. The complementation analysis (Example 6-3) showed that this ORF is the bioD gene.

30 The third ORF in the left gene cluster has a putative RBS upstream of the ATG initiation codon, and the nucleotide sequence of this gene is shown in SEQ ID NO: 3. This gene codes for a protein of 460 amino acid residues with a molecular weight of 51,731. The predicted amino acid sequence of this gene product is shown in SEQ ID NO: 4. This ORF was confirmed to correspond to the bioA gene (Example 6-3). An inverted repeat sequence was found to be

located approximately 3 bp downstream from the termination codon. This structure may act as a transcriptional terminator.

5 The first ORF in the right gene cluster, named ORF2 starts at the ATG codon preceded by a putative RBS. This gene product is a protein consisting of 63 amino acid residues, and the calculated molecular weight is 7,447. We could not identify the function of this gene product by the complementation analysis and the amino acid sequence homology search. Accordingly, this ORF was named ORF2.

10 The nucleotide sequence of the second ORF in the right gene cluster is shown in SEQ ID NO: 5. This gene has three potential ATG initiation codons corresponding to the first, twenty-fifth and thirty-second amino acid residues. The complementation analysis (Example 6-3) showed that this ORF corresponds to the bioF gene. The predicted amino acid sequence of this gene product is shown in SEQ ID NO: 6. The molecular weight of the predicted protein with  
15 387 amino acid residues was calculated to be 42,619, starting from the first initiation codon.

The third ORF in the right gene cluster as shown in SEQ ID NO: 7 has three potential initiation codons, two ATG codons (the first and eighteenth amino acid residues) and a GTG codon (the twelfth amino acid residue). The predicted amino acid sequence of this gene product  
20 is shown in SEQ ID NO: 8. The molecular weight of the predicted protein with 338 amino acid residues translated from the first initiation codon was calculated to be 37,438. The complementation analysis (Example 6-3) showed that this ORF corresponds to the bioB gene. The presence of an inverted repeat sequence 16 bp downstream from the termination codon is characteristic of a transcriptional terminator.

25 There were two possible promoter sequences forming face to face promoters between ORF1 and ORF2 as shown in Fig. 10. The transcriptions proceed to the left into the ORF1, bioD and bioA gene cluster, and to the right into the ORF2, bioF and bioB gene cluster. In addition, two transcriptional terminators were located downstream of the termination codons of  
30 the bioA and bioB genes. Therefore, the transcriptions in both directions generate two different mRNAs.

Two components of the inverted repeat sequences, Box1 and Box2, were found between the initiation site of the ORF1 and ORF2 genes (Fig. 10). The overall homology for the Box1  
35 and Box2 is 82.5%. Comparison of the Box1 or Box2 with the operator of the Escherichia coli

biotin operon [Nature, 276, 689-694, (1978)] showed that there is a high level of conservation (54.6% homology for both). The similarities between two inverted repeat sequences of the biotin operator of Escherichia coli suggest that the Box1 and Box2 must be involved in the negative control of the biotin synthesis by biotin.

5

#### Example 10

Nucleotide sequence of the inserted DNA fragments of pKH100.

10 The nucleotide sequence analysis of the inserted DNA fragment of pKH100 was performed in the same manner as described in Example 9. A gene cluster containing two ORFs was found on the inserted DNA fragment (Fig. 9-B). In addition, it was confirmed that a part of the vector plasmid pBR322 and the inserted DNA fragment were deleted.

15 The first ORF as shown in SEQ ID NO: 9 codes for a protein of 267 amino acid residues, and the calculated molecular weight is 29,423. The predicted amino acid sequence of this gene product is shown in SEQ ID NO: 10. A putative RBS is located at 6 bp upstream from the ATG initiation codon. The complementation analysis, as shown in Example 7, indicated that this ORF corresponds to the bioH gene.

20

The second ORF with a potential RBS was found downstream of the bioH gene. The ORF codes for a protein of 86 amino acid residues with a molecular weight of 9,955. The protein encoded by the ORF did not share homology with the biotin gene products of Escherichia coli and Bacillus sphaericus. The ORF was named ORF3.

25

A possible promoter sequence was found upstream from the initiation codon of the bioH gene as shown in Fig. 11. Since no inverted repeat sequence such as Box1 and Box2 was found in the 5'-noncoding region of the bioH gene, the transcription of this gene cluster must be not regulated. In addition, there is an inverted repeat sequence overlapping with the termination codon of ORF3. Since this structure is able to act as a transcriptional terminator, the putative bioH promoter would therefore allow transcription of the bioH and ORF3 genes.

30

### Example 11

Nucleotide sequence of the inserted DNA fragments of pKC100.

5        The nucleotide sequence analysis of the inserted DNA fragment of pKC100 was performed in the same manner as described in Example 9. A gene cluster consisting of three ORFs was found on the inserted DNA fragment (Fig. 9-C).

10        The third ORF has a putative RBS upstream of the initiation codon and the nucleotide sequence of this gene is shown in SEQ ID NO: 15. This gene codes for a protein of 276 amino acid residues, and the calculated molecular weight is 31,599. The predicted amino acid sequence of this gene product is shown in SEQ ID NO: 16. The complementation analysis as shown in Example 8 indicating that this ORF corresponds to the bioC gene.

15        The first ORF as shown in SEQ ID NO: 11 codes for a protein of 398 amino acid residues with a molecular weight of 44,776. A putative RBS is located upstream of the initiation codon. The predicted amino acid sequence of this gene product as shown in SEQ ID NO: 12 has 43.0% homology with that of the bioF gene product of *Kurthia* sp. 538-KA26 in Example 9. Moreover, the pKC100 complemented the *Escherichia coli* bioF mutant as shown in Example 4.  
20        Therefore, this ORF was concluded to be a gene for an isozyme of the bioF gene product, KAPA synthetase. Therefore, this ORF was named bioFII gene.

      The second ORF as shown in SEQ ID NO: 13 has a putative RBS upstream of the initiation codon. This gene codes for a protein of 248 amino acid residues with a molecular  
25        weight of 28,629. The predicted amino acid sequence of this gene product as shown in SEQ ID NO: 14 has 24.2% homology with that of the bioH gene product of *Kurthia* sp. 538-KA26 in Example 10. As shown in Example 4, the pKC100 also complemented the *Escherichia coli* bioH mutant. These results showed that this ORF is a gene for isozyme of the bioH gene product therefore this ORF was named bioHII gene.

30

      A possible promoter sequence was found upstream from the initiation codon of the bioFII gene as shown in Fig. 12. An inverted sequence is located between the promoter sequence and the RBS of the bioFII gene. This inverted repeat sequence designated Box3 was compared with the Box1 and Box2 located between the ORF1 and ORF2 genes (Example 9). The Box1, Box2  
35        and Box3 were extremely similar to each other (homology of Box1 and Box3 was 80.0% and

that of Box2 and Box3 was 77.5%). Therefore, the cluster of the bioC gene must be regulated by a negative control similarly to the bioA cluster and the bioB cluster. In addition, there is an inverted repeat sequence 254 bp downstream of the termination codon of the bioC gene. This structure is thought to act as a transcriptional terminator.

5

#### Example 12

Construction of the shuttle vector for *Escherichia coli* and *Kurthia* sp. strain.

10 A shuttle vector for *Escherichia coli* and *Kurthia* sp. was constructed by the strategy as shown in Fig. 13. The *Staphylococcus aureus* plasmid pUB110 (Bacillus Genetic Stock Center; The Ohio State University, Department of Biochemistry, 484 West Twelfth Avenue, Columbus, Ohio 43210, USA) was completely digested with EcoRI and PvuII. A 3.5 Kb EcoRI-PvuII fragment containing the replication origin for *Kurthia* sp. and the kanamycin resistant gene was  
15 isolated by agarose gel electrophoresis. The pUC19 was completely digested with EcoRI and DraI, and the 1.2 Kb EcoRI-DraI fragment having the replication origin of *Escherichia coli* was isolated by agarose gel electrophoresis. Then, these fragments were ligated with a DNA ligation Kit to generate the shuttle vector pYK1. pYK1 can replicate in *Escherichia coli* and *Kurthia* sp., and *Escherichia coli* or *Kurthia* sp. transformed by pYK1 show resistance to kanamycin.

20

#### Example 13

Construction of the expression plasmid of the bioB gene of *Kurthia* sp.

25 pYK114 in which the *Kurthia* bioB gene was inserted downstream of the promoter of the *Kurthia* bioH gene was constructed by the strategy as shown in Fig. 14. pKH101 of Example 7 was completely digested with BanII, and ends of BanII fragments were blunted by the Klenow fragment of the DNA polymerase. Then the BanII fragments were treated with EcoRI, and a 0.6 Kb EcoRI-blunt fragment containing the bioH promoter was isolated by agarose gel  
30 electrophoresis. pKB104 of Example 6 was completely digested with KpnI, and KpnI ends were changed to blunt ends by treatment with the Klenow fragment. After digestion with HindIII, a 1.3 Kb blunt-HindIII fragment carrying the bioB gene was isolated by agarose electrophoresis. The EcoRI-blunt and blunt-HindIII fragments were ligated with pYK1 digested with EcoRI and HindIII to construct pYK114. The bioB gene is constitutively expressed under  
35 the bioH promoter from pYK114.



#### Example 14

Isolation of the derivative strain of *Kurthia* sp. 538-51F9 with a high transformation efficiency.

*Kurthia* sp. 538-51F9 (DSM No.10610) was cultivated at 28°C in 50 ml of Trypticase Soy Broth (Becton Dickinson) until an optical density at 600 nm ( $OD_{600}$ ) of 1.0. Grown cells were collected by centrifugation and suspended in SMM (0.5 M sucrose, 0.02M sodium maleate, 0.02 M  $MgCl_2 \cdot 6H_2O$ ; pH 6.5) at  $OD_{600}$  16. Then lysozyme (Sigma) was added to the cell suspension at 200 mg/ml, and the suspension was incubated at 30°C for 90 minutes to form protoplasts. After the protoplasts have been washed with SMM twice, they were suspended in 0.5 ml of SMM. 1.5 ml of PEG solution (30% w/v polyethyleneglycol 4000 in SMM) was added to the protoplast suspension, and the suspension was incubated for 2 minutes on ice. Then 6 ml of SMM was added, and the protoplasts were collected by centrifugation. The collected protoplasts were suspended in SMM and incubated at 30°C for 90 minutes. DM3 medium (0.5 M sodium succinate pH 7.3, 0.5% w/v casamino acid, 0.5% w/v yeast extract, 0.3% w/v  $KH_2PO_4$ , 0.7% w/v  $K_2HPO_4$ , 0.5% w/v glucose, 0.02 M  $MgCl_2 \cdot 6H_2O$ , 0.01% w/v bovine serum albumin) containing 0.6% agarose (Sigma; Type VII) was added to the protoplast suspension, and the suspension was overlaid on DM3 medium agar plates. The plates were incubated at 30°C for 3 days. In total, 65 colonies regenerated on the DM3 plates were obtained.

The transformation efficiency of the regenerated strains was investigated with pYK1 of Example 12. As a result, 40 strains were selected and cultivated at 28°C in 50 ml of Trypticase Soy Broth until  $OD_{600}$  was 1.0. Grown cells were collected by centrifugation and suspended in SMM at  $OD_{600}$  16. Then the cells were treated with lysozyme by the method described above, and the protoplasts were obtained. The protoplasts were suspended in 0.5 ml SMM, and pYK1 (1 µg) was added to the protoplast suspensions. After addition of 1.5 ml of a PEG solution, the suspensions were incubated for 2 minutes on ice. 6 ml of SMM was added, and the protoplasts were collected by centrifugation. Then the protoplasts were suspended in SMM and incubated at 30°C for 90 minutes. The DM3 medium containing 0.6% agarose was added to the protoplast suspensions, and the suspensions were overlaid on DM3 medium agar plates. The plates were incubated at 30°C for 3 days. The DM3-agarose including the regenerated colonies on the plates were collected and spread on the nutrient broth agar plates with 5 µg/ml

kanamycin to select the transformants. The plates were incubated overnight at 30°C. Finally, the derivative strain, *Kurthia* sp. 538-51F9-RG21, characterized by a high transformation efficiency (2,000 transformants per µg of DNA) was obtained.

### Example 15

#### Amplification of the *bioB* gene in *Kurthia* sp. 538-51F9-RG21.

##### 1. Transformation of *Kurthia* sp. 538-51F9-RG21.

The expression plasmid of the *bioB* gene of the *Kurthia* strain, pYK114, was constructed as described in Example 13. *Kurthia* sp. 538-51F9-RG21 was transformed with pYK114 and the vector plasmid pYK1 as described in Example 14. *Kurthia* sp. 538-51F9-RG21 carrying pYK1 or pYK114 was named *Kurthia* sp. 538-51F9-RG21 (pYK1) or *Kurthia* sp. 538-51F9-RG21 (pYK114), respectively.

##### 2. Biotin production by fermentation.

*Kurthia* sp. 538-51F9-RG21 (pYK1) and *Kurthia* sp. 538-51F9-RG21 (pYK114) were inoculated into 50 ml of the production medium (6% glycerol, 5.5% proteose peptone, 0.1%  $\text{KH}_2\text{PO}_4$ , 0.05%  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.05%  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.001%  $\text{MnSO}_4 \cdot 5\text{H}_2\text{O}$ ; pH 7.0) containing 5 µg/ml kanamycin. As a control, *Kurthia* sp. 538-51F9-RG21 was inoculated into 50 ml of the production medium. The cultivation was carried out at 28°C for 120 hours.

After the cultivation, 2 ml of the culture broth was centrifuged to remove bacterial cells, and the supernatant was obtained. Biotin production in the supernatant was assayed by the microbiological assay using *Lactobacillus plantarum* (ATCC 8014). The amounts of produced biotin are given in Table 3.

Table 3

Strain of <i>Kurthia</i> sp.	Biotin production (mg/L)
51F9-RG21	15.4
51F9-RG21 (pYK1)	14.3
51F9-RG21 (pYK114)	39.0

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5

(i) APPLICANT: Furuichi, Yasuhiro  
Hoshino, Tatsuo  
Kimura, Hitoshi  
Kiyasu, Tatsuya  
Nagahashi, Yoshie

10

(ii) TITLE OF INVENTION: BIOTIN BIOSYNTHETIC GENES

15

(iii) NUMBER OF SEQUENCES: 16

(iv) CORRESPONDENCE ADDRESS:

20

(A) ADDRESSEE: Hoffmann-La Roche Inc.  
(B) STREET: 340 Kingsland Street  
(C) CITY: Nutley  
(D) STATE: NJ  
(E) COUNTRY: USA  
(F) ZIP: 07110

25

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

35

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 96115540.5  
(B) FILING DATE: 27-SEP-1996

40

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Pokras, Bruce A.  
(B) REGISTRATION NUMBER: 32,748  
(C) REFERENCE/DOCKET NUMBER: 4227/058

45

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (973) 235-5000  
(B) TELEFAX: (973) 235-2363

50

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 711 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Kurthia sp.

(B) STRAIN: 538-KA26

10

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..708

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG GGT CAA GCC TAC TTT ATA ACC GGA ACT GGC ACG GAT ATC GGA AAA 48  
Met Gly Gln Ala Tyr Phe Ile Thr Gly Thr Gly Thr Asp Ile Gly Lys  
1 5 10 15

20

ACC GTC GCC ACG AGT TTA CTC TAT ATG TCT CTT CAA ACA ATG GGA AAA 96  
Thr Val Ala Thr Ser Leu Leu Tyr Met Ser Leu Gln Thr Met Gly Lys  
20 25 30

25

AGC GTC ACA ATA TTT AAG CCG TTT CAA ACA GGA TTG ATT CAC GAA ACG 144  
Ser Val Thr Ile Phe Lys Pro Phe Gln Thr Gly Leu Ile His Glu Thr  
35 40 45

30

AAT ACA TAC CCT GAC ATC TCT TGG TTT GAG CAG GAA CTT GGT GTA AAG 192  
Asn Thr Tyr Pro Asp Ile Ser Trp Phe Glu Gln Glu Leu Gly Val Lys  
50 55 60

35

GCA CCT GGG TTT TAC ATG CTT GAA CCC GAA ACA TCT CCA CAC TTA GCT 240  
Ala Pro Gly Phe Tyr Met Leu Glu Pro Glu Thr Ser Pro His Leu Ala  
65 70 75 80

ATA AAA TTA ACA GGG CAA CAA ATC GAC GAG CAA AAG GTC GTG GAA CGA 288  
Ile Lys Leu Thr Gly Gln Gln Ile Asp Glu Gln Lys Val Val Glu Arg  
85 90 95

40

GTT CAC GAA CTC GAA CAA ATG TAT GAC ATC GTG TTA GTC GAG GGC GCT 336  
Val His Glu Leu Glu Gln Met Tyr Asp Ile Val Leu Val Glu Gly Ala  
100 105 110

45

GGG GGA TTG GCC GTA CCA CTC ATT GAA CGA GCG AAC AGT TTC TAT ATG 384  
Gly Gly Leu Ala Val Pro Leu Ile Glu Arg Ala Asn Ser Phe Tyr Met  
115 120 125

50

ACA ACC GAT TTA ATT AGA GAT TGC AAC ATG CCA GTC ATT TTC GTT TCT 432  
Thr Thr Asp Leu Ile Arg Asp Cys Asn Met Pro Val Ile Phe Val Ser  
130 135 140

55

ACA AGC GGT TTA GGA TCG ATT CAT AAT GTC ATA ACT ACG CAT TCG TAT 480  
Thr Ser Gly Leu Gly Ser Ile His Asn Val Ile Thr Thr His Ser Tyr  
145 150 155 160

GCC AAA TTG CAT GAT ATT AGC GTT AAA ACT ATT TTA TAT AAC CAT TAT 528  
Ala Lys Leu His Asp Ile Ser Val Lys Thr Ile Leu Tyr Asn His Tyr  
165 170 175

	CGG CCC GAC GAT GAA ATT CAT CGT GAC AAT ATC CTA ACC GTT GAA AAG	576
	Arg Pro Asp Asp Glu Ile His Arg Asp Asn Ile Leu Thr Val Glu Lys	
	180 185 190	
5	CTC ACA GGA CTC GCT GAC CTC GCC TGC ATA CCA ACA TTT GTC GAC GTA	624
	Leu Thr Gly Leu Ala Asp Leu Ala Cys Ile Pro Thr Phe Val Asp Val	
	195 200 205	
10	AGA AAA GAT CTG AGA GTC TAC ATA CTT GAT TTA CTT AGT AAT CAT GAA	672
	Arg Lys Asp Leu Arg Val Tyr Ile Leu Asp Leu Leu Ser Asn His Glu	
	210 215 220	
15	TTT ACT CAA CAA CTA AAA GAG GTG TTC AAG AAT GAA TAG	711
	Phe Thr Gln Gln Leu Lys Glu Val Phe Lys Asn Glu	
	225 230 235	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 236 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

30	Met Gly Gln Ala Tyr Phe Ile Thr Gly Thr Gly Thr Asp Ile Gly Lys
	1 5 10 15
	Thr Val Ala Thr Ser Leu Leu Tyr Met Ser Leu Gln Thr Met Gly Lys
	20 25 30
35	Ser Val Thr Ile Phe Lys Pro Phe Gln Thr Gly Leu Ile His Glu Thr
	35 40 45
40	Asn Thr Tyr Pro Asp Ile Ser Trp Phe Glu Gln Glu Leu Gly Val Lys
	50 55 60
	Ala Pro Gly Phe Tyr Met Leu Glu Pro Glu Thr Ser Pro His Leu Ala
	65 70 75 80
45	Ile Lys Leu Thr Gly Gln Gln Ile Asp Glu Gln Lys Val Val Glu Arg
	85 90 95
	Val His Glu Leu Glu Gln Met Tyr Asp Ile Val Leu Val Glu Gly Ala
	100 105 110
50	Gly Gly Leu Ala Val Pro Leu Ile Glu Arg Ala Asn Ser Phe Tyr Met
	115 120 125
55	Thr Thr Asp Leu Ile Arg Asp Cys Asn Met Pro Val Ile Phe Val Ser
	130 135 140
	Thr Ser Gly Leu Gly Ser Ile His Asn Val Ile Thr Thr His Ser Tyr
	145 150 155 160

Ala Lys Leu His Asp Ile Ser Val Lys Thr Ile Leu Tyr Asn His Tyr  
165 170 175

5 Arg Pro Asp Asp Glu Ile His Arg Asp Asn Ile Leu Thr Val Glu Lys  
180 185 190

Leu Thr Gly Leu Ala Asp Leu Ala Cys Ile Pro Thr Phe Val Asp Val  
195 200 205

10 Arg Lys Asp Leu Arg Val Tyr Ile Leu Asp Leu Leu Ser Asn His Glu  
210 215 220

15 Phe Thr Gln Gln Leu Lys Glu Val Phe Lys Asn Glu  
225 230 235

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
20 (A) LENGTH: 1383 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

30 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Kurthia sp.  
(B) STRAIN: 538-KA26

35 (ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..1380

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG AAT AGT CAT GAC TTA GAA AAG TGG GAT AAG GAA TAT GTA TGG CAT	48
Met Asn Ser His Asp Leu Glu Lys Trp Asp Lys Glu Tyr Val Trp His	
240 245 250	
45 CCG TTT ACA CAA ATG AAA ACG TAT CGA GAA AGT AAA CCG CTA ATC ATT	96
Pro Phe Thr Gln Met Lys Thr Tyr Arg Glu Ser Lys Pro Leu Ile Ile	
255 260 265	
50 GAA CGC GGG GAA GGG AGC TAC CTT TTT GAC ATA GAA GGC AAT CGG TAC	144
Glu Arg Gly Glu Gly Ser Tyr Leu Phe Asp Ile Glu Gly Asn Arg Tyr	
270 275 280	
55 TTG GAC GGT TAT GCT TCA TTA TGG GTC AAC GTA CAT GGC CAT AAT GAA	192
Leu Asp Gly Tyr Ala Ser Leu Trp Val Asn Val His Gly His Asn Glu	
285 290 295 300	

	CCA GAG CTA AAC AAC GCT CTC ATT GAA CAA GTT GAA AAA GTC GCA CAC Pro Glu Leu Asn Asn Ala Leu Ile Glu Gln Val Glu Lys Val Ala His	240
	305 310 315	
5	TCA ACA CTA CTA GGA TCT GCA AAT GTA CCA TCC ATA TTA CTG GCT AAG Ser Thr Leu Leu Gly Ser Ala Asn Val Pro Ser Ile Leu Leu Ala Lys	288
	320 325 330	
10	AAA TTA GCA GAG ATT ACT CCT GGT CAT TTA TCG AAA GTC TTT TAC TCG Lys Leu Ala Glu Ile Thr Pro Gly His Leu Ser Lys Val Phe Tyr Ser	336
	335 340 345	
15	GAC ACT GGA TCA GCT GCT GTA GAA ATC TCC CTT AAA GTC GCT TAT CAA Asp Thr Gly Ser Ala Ala Val Glu Ile Ser Leu Lys Val Ala Tyr Gln	384
	350 355 360	
20	TAT TGG AAA AAT ATC GAT CCT GTA AAG TAT CAA CAT AAA AAT AAA TTT Tyr Trp Lys Asn Ile Asp Pro Val Lys Tyr Gln His Lys Asn Lys Phe	432
	365 370 375 380	
	GTC TCC CTG AAC GAG GCG TAC CAC GGT GAT ACA GTT GGA GCA GTG AGT Val Ser Leu Asn Glu Ala Tyr His Gly Asp Thr Val Gly Ala Val Ser	480
	385 390 395	
25	GTC GGC GGA ATG GAT TTA TTC CAT AGA ATC TTT AAA CCA CTA CTA TTT Val Gly Gly Met Asp Leu Phe His Arg Ile Phe Lys Pro Leu Leu Phe	528
	400 405 410	
30	GAA CGG ATT CCA ACT CCT TCT CCT TAT ACA TAT CGC ATG GCT AAA CAC Glu Arg Ile Pro Thr Pro Ser Pro Tyr Thr Tyr Arg Met Ala Lys His	576
	415 420 425	
35	GGG GAT CAA GAA GCA GTG AAA AAC TAT TGT ATT GAT GAG CTG GAA AAG Gly Asp Gln Glu Ala Val Lys Asn Tyr Cys Ile Asp Glu Leu Glu Lys	624
	430 435 440	
40	TTG CTT CAA GAC CAA GCA GAG GAA ATT GCA GGA TTG ATT ATC GAA CCG Leu Leu Gln Asp Gln Ala Glu Glu Ile Ala Gly Leu Ile Ile Glu Pro	672
	445 450 455 460	
	CTT GTT CAA GGA GCA GCA GGC ATC ATT ACC CAC CCT CCT GGC TTT TTA Leu Val Gln Gly Ala Ala Gly Ile Ile Thr His Pro Pro Gly Phe Leu	720
	465 470 475	
45	AAA GCG GTC GAA CAA TTG TGC AAG AAG TAC AAT ATA TTA TTG ATT TGT Lys Ala Val Glu Gln Leu Cys Lys Lys Tyr Asn Ile Leu Leu Ile Cys	768
	480 485 490	
50	GAC GAA GTA GCG GTA GGA TTT GGT CGC ACC GGT ACA TTA TTT GCC TGT Asp Glu Val Ala Val Gly Phe Gly Arg Thr Gly Thr Leu Phe Ala Cys	816
	495 500 505	
55	GAA CAA GAA GAT GTC GTC CCT GAT ATT ATG TGT ATC GGT AAA GGA ATT Glu Gln Glu Asp Val Val Pro Asp Ile Met Cys Ile Gly Lys Gly Ile	864
	510 515 520	

	ACT GGC GGC TAT ATG CCT CTG GCG GCC ACT ATC ATG AAC GAA CAA ATC	912
	Thr Gly Gly Tyr Met Pro Leu Ala Ala Thr Ile Met Asn Glu Gln Ile	
	525 530 535 540	
5	TTT AAT TCT TTT TTA GGA GAG CCC GAT GAA CAT AAA ACC TTC TAT CAC	960
	Phe Asn Ser Phe Leu Gly Glu Pro Asp Glu His Lys Thr Phe Tyr His	
	545 550 555	
10	GGC CAC ACC TAC ACA GGG AAT CAA CTA GCC TGT GCC CTG GCG CTG AAG	1008
	Gly His Thr Tyr Thr Gly Asn Gln Leu Ala Cys Ala Leu Ala Leu Lys	
	560 565 570	
15	AAT ATC GAA CTA ATA GAA AGA CGA GAT CTC GTC AAA GAC ATC CAG AAG	1056
	Asn Ile Glu Leu Ile Glu Arg Arg Asp Leu Val Lys Asp Ile Gln Lys	
	575 580 585	
20	AAA TCC AAG CAG CTA TCT GAA AAA CTG CAA TCG CTA TAT GAA CTC CCG	1104
	Lys Ser Lys Gln Leu Ser Glu Lys Leu Gln Ser Leu Tyr Glu Leu Pro	
	590 595 600	
25	ATT GTC GGT GAT ATC CGC CAG CGC GGC CTC ATG ATT GGA ATA GAA ATC	1152
	Ile Val Gly Asp Ile Arg Gln Arg Gly Leu Met Ile Gly Ile Glu Ile	
	605 610 615 620	
30	GTT AAA GAT CGC CAA ACA AAA GAA CCG TTC ACA ATC CAA GAA AAT ATC	1200
	Val Lys Asp Arg Gln Thr Lys Glu Pro Phe Thr Ile Gln Glu Asn Ile	
	625 630 635	
35	GTT TCA AGC ATC ATC CAA AAC GCT CGG GAA AAT GGC CTG ATC ATT CGG	1248
	Val Ser Ser Ile Ile Gln Asn Ala Arg Glu Asn Gly Leu Ile Ile Arg	
	640 645 650	
40	GAA CTT GGC CCT GTC ATC ACA ATG ATG CCC ATT CTT TCC ATG TCA GAA	1296
	Glu Leu Gly Pro Val Ile Thr Met Met Pro Ile Leu Ser Met Ser Glu	
	655 660 665	
45	AAG GAA CTG AAT ACT ATG GTC GAA ACT GTC TAC CGT TCG ATA CAG GAC	1344
	Lys Glu Leu Asn Thr Met Val Glu Thr Val Tyr Arg Ser Ile Gln Asp	
	670 675 680	
50	GTT TCT GTG CAC AAC GGA TTA ATC CCA GCA GCA AAC TGA	1383
	Val Ser Val His Asn Gly Leu Ile Pro Ala Ala Asn	
	685 690 695	

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 460 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Asn	Ser	His	Asp	Leu	Glu	Lys	Trp	Asp	Lys	Glu	Tyr	Val	Trp	His
1					5				10					15	



Pro Phe Thr Gln Met Lys Thr Tyr Arg Glu Ser Lys Pro Leu Ile Ile  
 20 25 30  
 5 Glu Arg Gly Glu Gly Ser Tyr Leu Phe Asp Ile Glu Gly Asn Arg Tyr  
 35 40 45  
 Leu Asp Gly Tyr Ala Ser Leu Trp Val Asn Val His Gly His Asn Glu  
 50 55 60  
 10 Pro Glu Leu Asn Asn Ala Leu Ile Glu Gln Val Glu Lys Val Ala His  
 65 70 75 80  
 Ser Thr Leu Leu Gly Ser Ala Asn Val Pro Ser Ile Leu Leu Ala Lys  
 85 90 95  
 Lys Leu Ala Glu Ile Thr Pro Gly His Leu Ser Lys Val Phe Tyr Ser  
 100 105 110  
 20 Asp Thr Gly Ser Ala Ala Val Glu Ile Ser Leu Lys Val Ala Tyr Gln  
 115 120 125  
 Tyr Trp Lys Asn Ile Asp Pro Val Lys Tyr Gln His Lys Asn Lys Phe  
 130 135 140  
 25 Val Ser Leu Asn Glu Ala Tyr His Gly Asp Thr Val Gly Ala Val Ser  
 145 150 155 160  
 Val Gly Gly Met Asp Leu Phe His Arg Ile Phe Lys Pro Leu Leu Phe  
 165 170 175  
 Glu Arg Ile Pro Thr Pro Ser Pro Tyr Thr Tyr Arg Met Ala Lys His  
 180 185 190  
 35 Gly Asp Gln Glu Ala Val Lys Asn Tyr Cys Ile Asp Glu Leu Glu Lys  
 195 200 205  
 Leu Leu Gln Asp Gln Ala Glu Glu Ile Ala Gly Leu Ile Ile Glu Pro  
 210 215 220  
 40 Leu Val Gln Gly Ala Ala Gly Ile Ile Thr His Pro Pro Gly Phe Leu  
 225 230 235 240  
 Lys Ala Val Glu Gln Leu Cys Lys Lys Tyr Asn Ile Leu Leu Ile Cys  
 245 250 255  
 45 Asp Glu Val Ala Val Gly Phe Gly Arg Thr Gly Thr Leu Phe Ala Cys  
 260 265 270  
 50 Glu Gln Glu Asp Val Val Pro Asp Ile Met Cys Ile Gly Lys Gly Ile  
 275 280 285  
 Thr Gly Gly Tyr Met Pro Leu Ala Ala Thr Ile Met Asn Glu Gln Ile  
 290 295 300  
 55 Phe Asn Ser Phe Leu Gly Glu Pro Asp Glu His Lys Thr Phe Tyr His  
 305 310 315 320

Gly His Thr Tyr Thr Gly Asn Gln Leu Ala Cys Ala Leu Ala Leu Lys  
 325 330 335  
 5 Asn Ile Glu Leu Ile Glu Arg Arg Asp Leu Val Lys Asp Ile Gln Lys  
 340 345 350  
 Lys Ser Lys Gln Leu Ser Glu Lys Leu Gln Ser Leu Tyr Glu Leu Pro  
 355 360 365  
 10 Ile Val Gly Asp Ile Arg Gln Arg Gly Leu Met Ile Gly Ile Glu Ile  
 370 375 380  
 Val Lys Asp Arg Gln Thr Lys Glu Pro Phe Thr Ile Gln Glu Asn Ile  
 385 390 395 400  
 15 Val Ser Ser Ile Ile Gln Asn Ala Arg Glu Asn Gly Leu Ile Ile Arg  
 405 410 415  
 20 Glu Leu Gly Pro Val Ile Thr Met Met Pro Ile Leu Ser Met Ser Glu  
 420 425 430  
 Lys Glu Leu Asn Thr Met Val Glu Thr Val Tyr Arg Ser Ile Gln Asp  
 435 440 445  
 25 Val Ser Val His Asn Gly Leu Ile Pro Ala Ala Asn  
 450 455 460

(2) INFORMATION FOR SEQ ID NO:5:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1164 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

40 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Kurthia sp.

(B) STRAIN: 538-KA26

45

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1161

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

55 ATG ATT TGG GAG AAG GAA CTA GAA AAG ATT AAA GAA GGA GGG CTT TAC  
 Met Ile Trp Glu Lys Glu Leu Glu Lys Ile Lys Glu Gly Gly Leu Tyr  
 465 470 475

48

	AGA CAA CTC CAA ACC GTT GAA ACA ATG AGC GAT CAA GGG TAT GCC ATG Arg Gln Leu Gln Thr Val Glu Thr Met Ser Asp Gln Gly Tyr Ala Met 480 485 490	96
5	GTG AAC GGA AAA AAA ATG ATG ATG TTT GCC TCC AAT AAT TAC TTA GGG Val Asn Gly Lys Lys Met Met Met Phe Ala Ser Asn Asn Tyr Leu Gly 495 500 505	144
10	ATT GCC AAT GAT CAA CGA TTA ATT GAG GCT TCT GTC CAA GCG ACT CAA Ile Ala Asn Asp Gln Arg Leu Ile Glu Ala Ser Val Gln Ala Thr Gln 510 515 520	192
15	AGA TTT GGT ACG GGT TCT ACT GGT TCA CGA TTA ACC ACT GGC AAT ACA Arg Phe Gly Thr Gly Ser Thr Gly Ser Arg Leu Thr Thr Gly Asn Thr 525 530 535 540	240
20	ATT GTC CAT GAA AAA CTA GAG AAA AGA CTT GCA GAG TTT AAG CAA ACG Ile Val His Glu Lys Leu Glu Lys Arg Leu Ala Glu Phe Lys Gln Thr 545 550 555	288
	GAT GCA GCG ATA GTA TTA AAC ACA GGG TAT ATG GCT AAC ATA GCA GCG Asp Ala Ala Ile Val Leu Asn Thr Gly Tyr Met Ala Asn Ile Ala Ala 560 565 570	336
25	TTA ACG ACC CTT GTT GGT AGT GAC GAT CTC ATT TTA TCC GAT GAG ATG Leu Thr Thr Leu Val Gly Ser Asp Asp Leu Ile Leu Ser Asp Glu Met 575 580 585	384
30	AAT CAT GCC AGT ATT ATT GAT GGC TGC CGT TTA TCA CGT GCG GAA ACT Asn His Ala Ser Ile Ile Asp Gly Cys Arg Leu Ser Arg Ala Glu Thr 590 595 600	432
35	ATC ATT TAT CGT CAT GCT GAT TTA CTT GAC TTG GAA ATG AAA CTC CAG Ile Ile Tyr Arg His Ala Asp Leu Leu Asp Leu Glu Met Lys Leu Gln 605 610 615 620	480
40	ATT AAT ACC CGC TAC AGG AAA AGA ATA ATT GTA ACG GAT GGC GTC TTT Ile Asn Thr Arg Tyr Arg Lys Arg Ile Ile Val Thr Asp Gly Val Phe 625 630 635	528
	TCG ATG GAT GGT GAT ATT GCG CCA TTG CCA GGT ATT GTC GAA CTT GCC Ser Met Asp Gly Asp Ile Ala Pro Leu Pro Gly Ile Val Glu Leu Ala 640 645 650	576
45	AAG CGT TAT GAT GCA CTT GTT ATG GTG GAT GAC GCA CAT GCG ACG GGT Lys Arg Tyr Asp Ala Leu Val Met Val Asp Asp Ala His Ala Thr Gly 655 660 665	624
50	GTT TTA GGT AAA GAC GGA AGG GGA ACT TCT GAA CAT TTT GGA CTG AAG Val Leu Gly Lys Asp Gly Arg Gly Thr Ser Glu His Phe Gly Leu Lys 670 675 680	672
55	GGG AAG ATA GAT ATC GAG ATG GGG ACA CTC TCC AAA GCT GTT GGT GCA Gly Lys Ile Asp Ile Glu Met Gly Thr Leu Ser Lys Ala Val Gly Ala 685 690 695 700	720

	GAA GGA GGG TAT ATC GCT GGA AGC AGG TCT TTA GTT GAC TAT GTC TTA	768
	Glu Gly Gly Tyr Ile Ala Gly Ser Arg Ser Leu Val Asp Tyr Val Leu	
	705 710 715	
5	AAT CGA GCC AGA CCG TTT GTC TTC TCT ACC GCC TTA TCA GCA GGA GTA	816
	Asn Arg Ala Arg Pro Phe Val Phe Ser Thr Ala Leu Ser Ala Gly Val	
	720 725 730	
10	GTA GCA AGT GCA CTT ACA GCA GTC GAT ATC ATT CAA TCA GAA CCT GAA	864
	Val Ala Ser Ala Leu Thr Ala Val Asp Ile Ile Gln Ser Glu Pro Glu	
	735 740 745	
15	CGC AGA GTA CGC ATT CGA GCC ATG AGC CAG CGT CTT TAT AAT GAA TTA	912
	Arg Arg Val Arg Ile Arg Ala Met Ser Gln Arg Leu Tyr Asn Glu Leu	
	750 755 760	
20	ACC TCC CTT GGC TAC ACA GTT TCG GGG GGA GAA ACT CCG ATT CTT GCC	960
	Thr Ser Leu Gly Tyr Thr Val Ser Gly Gly Glu Thr Pro Ile Leu Ala	
	765 770 775 780	
25	ATT ATT TGC GGA GAA CCG GAA CAG GCC ATG TTC CTT TCG AAA GAA TTA	1008
	Ile Ile Cys Gly Glu Pro Glu Gln Ala Met Phe Leu Ser Lys Glu Leu	
	785 790 795	
30	CAT AAG CAC GGA ATT TAT GCA CCA GCT ATC CGT TCG CCA ACG GTA CCT	1056
	His Lys His Gly Ile Tyr Ala Pro Ala Ile Arg Ser Pro Thr Val Pro	
	800 805 810	
35	CTT GGA ACT TCG CGC ATT CGA CTT ACG TTA ATG GCG ACA CAT CAA GAA	1104
	Leu Gly Thr Ser Arg Ile Arg Leu Thr Leu Met Ala Thr His Gln Glu	
	815 820 825	
40	GAA CAA ATG AAT CAT GTT ATC GAC GTG TTC AGA ACA ATC CTT ACC AAT	1152
	Glu Gln Met Asn His Val Ile Asp Val Phe Arg Thr Ile Leu Thr Asn	
	830 835 840	
45	AGA TAC AAA TAG	1164
	Arg Tyr Lys	
	845	

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 387 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Ile	Trp	Glu	Lys	Glu	Leu	Glu	Lys	Ile	Lys	Glu	Gly	Gly	Leu	Tyr
1				5				10						15	
Arg	Gln	Leu	Gln	Thr	Val	Glu	Thr	Met	Ser	Asp	Gln	Gly	Tyr	Ala	Met
			20					25					30		

Val Asn Gly Lys Lys Met Met Met Phe Ala Ser Asn Asn Tyr Leu Gly  
 35 40 45  
 5 Ile Ala Asn Asp Gln Arg Leu Ile Glu Ala Ser Val Gln Ala Thr Gln  
 50 55 60  
 Arg Phe Gly Thr Gly Ser Thr Gly Ser Arg Leu Thr Thr Gly Asn Thr  
 65 70 75 80  
 10 Ile Val His Glu Lys Leu Glu Lys Arg Leu Ala Glu Phe Lys Gln Thr  
 85 90 95  
 Asp Ala Ala Ile Val Leu Asn Thr Gly Tyr Met Ala Asn Ile Ala Ala  
 100 105 110  
 15 Leu Thr Thr Leu Val Gly Ser Asp Asp Leu Ile Leu Ser Asp Glu Met  
 115 120 125  
 20 Asn His Ala Ser Ile Ile Asp Gly Cys Arg Leu Ser Arg Ala Glu Thr  
 130 135 140  
 Ile Ile Tyr Arg His Ala Asp Leu Leu Asp Leu Glu Met Lys Leu Gln  
 145 150 155 160  
 25 Ile Asn Thr Arg Tyr Arg Lys Arg Ile Ile Val Thr Asp Gly Val Phe  
 165 170 175  
 Ser Met Asp Gly Asp Ile Ala Pro Leu Pro Gly Ile Val Glu Leu Ala  
 180 185 190  
 30 Lys Arg Tyr Asp Ala Leu Val Met Val Asp Asp Ala His Ala Thr Gly  
 195 200 205  
 35 Val Leu Gly Lys Asp Gly Arg Gly Thr Ser Glu His Phe Gly Leu Lys  
 210 215 220  
 Gly Lys Ile Asp Ile Glu Met Gly Thr Leu Ser Lys Ala Val Gly Ala  
 225 230 235 240  
 40 Glu Gly Gly Tyr Ile Ala Gly Ser Arg Ser Leu Val Asp Tyr Val Leu  
 245 250 255  
 Asn Arg Ala Arg Pro Phe Val Phe Ser Thr Ala Leu Ser Ala Gly Val  
 260 265 270  
 45 Val Ala Ser Ala Leu Thr Ala Val Asp Ile Ile Gln Ser Glu Pro Glu  
 275 280 285  
 Arg Arg Val Arg Ile Arg Ala Met Ser Gln Arg Leu Tyr Asn Glu Leu  
 290 295 300  
 Thr Ser Leu Gly Tyr Thr Val Ser Gly Gly Glu Thr Pro Ile Leu Ala  
 305 310 315 320  
 55 Ile Ile Cys Gly Glu Pro Glu Gln Ala Met Phe Leu Ser Lys Glu Leu  
 325 330 335

His Lys His Gly Ile Tyr Ala Pro Ala Ile Arg Ser Pro Thr Val Pro  
 340 345 350  
 5 Leu Gly Thr Ser Arg Ile Arg Leu Thr Leu Met Ala Thr His Gln Glu  
 355 360 365  
 Glu Gln Met Asn His Val Ile Asp Val Phe Arg Thr Ile Leu Thr Asn  
 370 375 380  
 10 Arg Tyr Lys  
 385  
 (2) INFORMATION FOR SEQ ID NO:7:  
 15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1017 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 20 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 25 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Kurthia sp.  
 (B) STRAIN: 538-KA26  
 30 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1014  
 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:  
 ATG AGA AAA GAG GGA TTA GGT TTG GAA ACA TTG GTG AAA AAG GAT TGG 48  
 40 Met Arg Lys Glu Gly Leu Gly Leu Glu Thr Leu Val Lys Lys Asp Trp  
 390 395 400  
 AAG ATG CTA GCG GAA AAC GTA ATC AAA GGA TAT AAA GTA ACA GCG GAA 96  
 Lys Met Leu Ala Glu Asn Val Ile Lys Gly Tyr Lys Val Thr Ala Glu  
 405 410 415  
 45 GAA GCA CTT GCT ATT GTA CAA GCA CCT GAC AAC GAG GTT TTA GAG ATT 144  
 Glu Ala Leu Ala Ile Val Gln Ala Pro Asp Asn Glu Val Leu Glu Ile  
 420 425 430 435  
 50 TTG AAT GCA GCT TTC CTT ATT CGT CAG CAC TAT TAT GGA AAA AAG GTT 192  
 Leu Asn Ala Ala Phe Leu Ile Arg Gln His Tyr Tyr Gly Lys Lys Val  
 440 445 450  
 AAA TTG AAT ATG ATC ATT AAT ACG AAG TCA GGT CTA TGT CCT GAA GAT 240  
 55 Lys Leu Asn Met Ile Ile Asn Thr Lys Ser Gly Leu Cys Pro Glu Asp  
 455 460 465

	TGT GGC TAT TGT TCG CAG TCA ATC GTG TCG GAA GCT CCT ATC GAT AAA Cys Gly Tyr Cys Ser Gln Ser Ile Val Ser Glu Ala Pro Ile Asp Lys 470 475 480	288
5	TAT GCT TGG CTG ACC AAA GAG AAG ATT GTT GAA GGT GCT CAA GAA TCA Tyr Ala Trp Leu Thr Lys Glu Lys Ile Val Glu Gly Ala Gln Glu Ser 485 490 495	336
10	ATT CGT CGC AAA GCT GGC ACG TAT TGT ATC GTT GCT TCT GGC CGT CGT Ile Arg Arg Lys Ala Gly Thr Tyr Cys Ile Val Ala Ser Gly Arg Arg 500 505 510 515	384
15	CCG ACC AAT AGG GAA ATT GAT CAT GTC ATT GAA GCT GTG AAA GAA ATT Pro Thr Asn Arg Glu Ile Asp His Val Ile Glu Ala Val Lys Glu Ile 520 525 530	432
20	CGC GAG ACA ACG GAT CTT AAA ATA TGC TGC TGT CTA GGT TTC TTA AAT Arg Glu Thr Thr Asp Leu Lys Ile Cys Cys Cys Leu Gly Phe Leu Asn 535 540 545	480
	GAA ACG CAT GCC AGT AAG CTA GCT GAA GCT GGG GTT CAT CGC TAC AAG Glu Thr His Ala Ser Lys Leu Ala Glu Ala Gly Val His Arg Tyr Lys 550 555 560	528
25	CAC AAC TTA AAT ACA TCT CAA GAT AAT TAT AAG AAT ATT ACA TCC ACA His Asn Leu Asn Thr Ser Gln Asp Asn Tyr Lys Asn Ile Thr Ser Thr 565 570 575	576
30	CAT ACT TAT GAG GAC CGT GTA GAT ACA GTC GAA GCT GTA AAA GAG GCC His Thr Tyr Glu Asp Arg Val Asp Thr Val Glu Ala Val Lys Glu Ala 580 585 590 595	624
35	GGA ATG TCT CCA TGC TCG GGT GCC ATT TTT GGT ATG AAT GAG TCT AAT Gly Met Ser Pro Cys Ser Gly Ala Ile Phe Gly Met Asn Glu Ser Asn 600 605 610	672
40	GAA GAA GCA GTA GAG ATT GCC CTA TCC CTA CGC AGT CTT GAC GCG GAT Glu Glu Ala Val Glu Ile Ala Leu Ser Leu Arg Ser Leu Asp Ala Asp 615 620 625	720
	TCT ATT CCT TGT AAT TTC CTC AAT GCG ATT GAC GGT ACA CCA CTT GAG Ser Ile Pro Cys Asn Phe Leu Asn Ala Ile Asp Gly Thr Pro Leu Glu 630 635 640	768
45	GGA ACT TCC GAG TTG ACT CCA ACT AAA TGT TTG AAA TTA ATT TCG ATG Gly Thr Ser Glu Leu Thr Pro Thr Lys Cys Leu Lys Leu Ile Ser Met 645 650 655	816
50	ATG AGA TTT GTT AAT CCA AGT AAG GAA ATC CGT CTT GCT GGA GGT CGC Met Arg Phe Val Asn Pro Ser Lys Glu Ile Arg Leu Ala Gly Gly Arg 660 665 670 675	864
55	GAG GTG AAC CTC CGT TCC ATG CAA CCC ATG GCA CTT TAT GCA GCC AAT Glu Val Asn Leu Arg Ser Met Gln Pro Met Ala Leu Tyr Ala Ala Asn 680 685 690	912

TCT ATC TTC GTC GGC GAT TAT CTA ACA ACA GCT GGA CAA GAA CCT ACG 960  
 Ser Ile Phe Val Gly Asp Tyr Leu Thr Thr Ala Gly Gln Glu Pro Thr  
 695 700 705  
 5 GCG GAT TGG GGC ATT ATC GAA GAC CTT GGT TTT GAA ATT GAA GAA TGC 1008  
 Ala Asp Trp Gly Ile Ile Glu Asp Leu Gly Phe Glu Ile Glu Glu Cys  
 710 715 720  
 10 GCT CTT TAA 1017  
 Ala Leu  
 725

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 338 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

25 Met Arg Lys Glu Gly Leu Gly Leu Glu Thr Leu Val Lys Lys Asp Trp  
 1 5 10 15  
 Lys Met Leu Ala Glu Asn Val Ile Lys Gly Tyr Lys Val Thr Ala Glu  
 20 25 30  
 30 Glu Ala Leu Ala Ile Val Gln Ala Pro Asp Asn Glu Val Leu Glu Ile  
 35 40 45  
 Leu Asn Ala Ala Phe Leu Ile Arg Gln His Tyr Tyr Gly Lys Lys Val  
 50 55 60  
 Lys Leu Asn Met Ile Ile Asn Thr Lys Ser Gly Leu Cys Pro Glu Asp  
 65 70 75 80  
 40 Cys Gly Tyr Cys Ser Gln Ser Ile Val Ser Glu Ala Pro Ile Asp Lys  
 85 90 95  
 Tyr Ala Trp Leu Thr Lys Glu Lys Ile Val Glu Gly Ala Gln Glu Ser  
 100 105 110  
 45 Ile Arg Arg Lys Ala Gly Thr Tyr Cys Ile Val Ala Ser Gly Arg Arg  
 115 120 125  
 Pro Thr Asn Arg Glu Ile Asp His Val Ile Glu Ala Val Lys Glu Ile  
 130 135 140  
 Arg Glu Thr Thr Asp Leu Lys Ile Cys Cys Cys Leu Gly Phe Leu Asn  
 145 150 155 160  
 55 Glu Thr His Ala Ser Lys Leu Ala Glu Ala Gly Val His Arg Tyr Lys  
 165 170 175



5 His Asn Leu Asn Thr Ser Gln Asp Asn Tyr Lys Asn Ile Thr Ser Thr  
 180 185 190  
 His Thr Tyr Glu Asp Arg Val Asp Thr Val Glu Ala Val Lys Glu Ala  
 195 200 205  
 Gly Met Ser Pro Cys Ser Gly Ala Ile Phe Gly Met Asn Glu Ser Asn  
 210 215 220  
 10 Glu Glu Ala Val Glu Ile Ala Leu Ser Leu Arg Ser Leu Asp Ala Asp  
 225 230 235 240  
 Ser Ile Pro Cys Asn Phe Leu Asn Ala Ile Asp Gly Thr Pro Leu Glu  
 245 250 255  
 15 Gly Thr Ser Glu Leu Thr Pro Thr Lys Cys Leu Lys Leu Ile Ser Met  
 260 265 270  
 20 Met Arg Phe Val Asn Pro Ser Lys Glu Ile Arg Leu Ala Gly Gly Arg  
 275 280 285  
 Glu Val Asn Leu Arg Ser Met Gln Pro Met Ala Leu Tyr Ala Ala Asn  
 290 295 300  
 25 Ser Ile Phe Val Gly Asp Tyr Leu Thr Thr Ala Gly Gln Glu Pro Thr  
 305 310 315 320  
 Ala Asp Trp Gly Ile Ile Glu Asp Leu Gly Phe Glu Ile Glu Glu Cys  
 325 330 335  
 30 Ala Leu

(2) INFORMATION FOR SEQ ID NO:9:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 804 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 45 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
 50 (A) ORGANISM: Kurthia sp.  
 (B) STRAIN: 538-KA26  
 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 55 (B) LOCATION: 1..801

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

5	ATG CCA TTC GTA AAT CAT GAC AAT GAA AGC CTT TAC TAT GAG GTT CAC Met Pro Phe Val Asn His Asp Asn Glu Ser Leu Tyr Tyr Glu Val His 340 345 350	48
10	GGA CAA GGT GAT CCT TTA TTG TTG ATT ATG GGG CTC GGC TAT AAC TCT Gly Gln Gly Asp Pro Leu Leu Ile Met Gly Leu Gly Tyr Asn Ser 355 360 365 370	96
15	TTA TCC TGG CAT AGA ACG GTG CCC ACT TTA GCT AAG CGC TTT AAA GTA Leu Ser Trp His Arg Thr Val Pro Thr Leu Ala Lys Arg Phe Lys Val 375 380 385	144
20	ATC GTT TTT GAT AAT CGT GGT GTT GGT AAG AGC AGT AAG CCT GAA CAG Ile Val Phe Asp Asn Arg Gly Val Gly Lys Ser Ser Lys Pro Glu Gln 390 395 400	192
25	CCA TAT TCT ATT GAA ATG ATG GCT GAG GAT GCA AGA GCG GTC CTT GAT Pro Tyr Ser Ile Glu Met Met Ala Glu Asp Ala Arg Ala Val Leu Asp 405 410 415	240
30	GCT GTT TCG GTT GAC TCA GCA CAT GTA TAT GGG ATT TCA ATG GGT GGA Ala Val Ser Val Asp Ser Ala His Val Tyr Gly Ile Ser Met Gly Gly 420 425 430	288
35	ATG ATT GCC CAA AGG CTG GCA ATC ACA TAT CCA GAA CGT GTT CGT TCT Met Ile Ala Gln Arg Leu Ala Ile Thr Tyr Pro Glu Arg Val Arg Ser 435 440 445 450	336
40	CTT GTT CTA GGT TGT ACC ACT GCG GGT GGT ACT ACT CAT ATT CAA CCT Leu Val Leu Gly Cys Thr Thr Ala Gly Gly Thr Thr His Ile Gln Pro 455 460 465	384
45	TCT CCA GAA ATA TCT ACT TTA ATG GTA TCT CGA GCC TCC CTT ACA GGT Ser Pro Glu Ile Ser Thr Leu Met Val Ser Arg Ala Ser Leu Thr Gly 470 475 480	432
50	TCT CCA AGG GAT AAT GCC TGG TTA GCG GCA CCA ATA GTT TAT AGT CAA Ser Pro Arg Asp Asn Ala Trp Leu Ala Ala Pro Ile Val Tyr Ser Gln 485 490 495	480
55	GCT TTT ATT GAG AAG CAC CCT GAA TTA ATT CAG GAA GAT ATC CAA AAG Ala Phe Ile Glu Lys His Pro Glu Leu Ile Gln Glu Asp Ile Gln Lys 500 505 510	528
60	CGA ATA GAA ATC ATT ACT CCG CCA AGC GCC TAT CTG TCT CAA CTA CAA Arg Ile Glu Ile Ile Thr Pro Pro Ser Ala Tyr Leu Ser Gln Leu Gln 515 520 525 530	576
65	GCT TGT CTA ACT CAT GAT ACA TCC AAT GAA CTT GAT AAA ATA AAC ATA Ala Cys Leu Thr His Asp Thr Ser Asn Glu Leu Asp Lys Ile Asn Ile 535 540 545	624
70	CCA ACA TTG ATT ATA CAC GGT GAT GCA GAT AAT TTG GTT CCA TAT GAA Pro Thr Leu Ile Ile His Gly Asp Ala Asp Asn Leu Val Pro Tyr Glu 550 555 560	672

	AAC GGT AAA ATG TTA GCT GAA CGT ATT CAG GGT TCT CAG TTT CAC ACC	720
	Asn Gly Lys Met Leu Ala Glu Arg Ile Gln Gly Ser Gln Phe His Thr	
	565 570 575	
5	GTA TCC TGT GCT GGC CAC ATT TAT TTA ACA GAA GCA GCT AAG GAA GCA	768
	Val Ser Cys Ala Gly His Ile Tyr Leu Thr Glu Ala Ala Lys Glu Ala	
	580 585 590	
10	AAT GAC AAA GTT ATA CAG TTT CTA GCT CAT CTA TAA	804
	Asn Asp Lys Val Ile Gln Phe Leu Ala His Leu	
	595 600 605	

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 267 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

25	Met Pro Phe Val Asn His Asp Asn Glu Ser Leu Tyr Tyr Glu Val His
	1 5 10 15
	Gly Gln Gly Asp Pro Leu Leu Leu Ile Met Gly Leu Gly Tyr Asn Ser
	20 25 30
30	Leu Ser Trp His Arg Thr Val Pro Thr Leu Ala Lys Arg Phe Lys Val
	35 40 45
	Ile Val Phe Asp Asn Arg Gly Val Gly Lys Ser Ser Lys Pro Glu Gln
35	50 55 60
	Pro Tyr Ser Ile Glu Met Met Ala Glu Asp Ala Arg Ala Val Leu Asp
	65 70 75 80
40	Ala Val Ser Val Asp Ser Ala His Val Tyr Gly Ile Ser Met Gly Gly
	85 90 95
	Met Ile Ala Gln Arg Leu Ala Ile Thr Tyr Pro Glu Arg Val Arg Ser
	100 105 110
45	Leu Val Leu Gly Cys Thr Thr Ala Gly Gly Thr Thr His Ile Gln Pro
	115 120 125
	Ser Pro Glu Ile Ser Thr Leu Met Val Ser Arg Ala Ser Leu Thr Gly
50	130 135 140
	Ser Pro Arg Asp Asn Ala Trp Leu Ala Ala Pro Ile Val Tyr Ser Gln
	145 150 155 160
55	Ala Phe Ile Glu Lys His Pro Glu Leu Ile Gln Glu Asp Ile Gln Lys
	165 170 175

Arg Ile Glu Ile Ile Thr Pro Pro Ser Ala Tyr Leu Ser Gln Leu Gln  
 180 185 190  
 5 Ala Cys Leu Thr His Asp Thr Ser Asn Glu Leu Asp Lys Ile Asn Ile  
 195 200 205  
 Pro Thr Leu Ile Ile His Gly Asp Ala Asp Asn Leu Val Pro Tyr Glu  
 210 215 220  
 10 Asn Gly Lys Met Leu Ala Glu Arg Ile Gln Gly Ser Gln Phe His Thr  
 225 230 235 240  
 Val Ser Cys Ala Gly His Ile Tyr Leu Thr Glu Ala Ala Lys Glu Ala  
 245 250 255  
 15 Asn Asp Lys Val Ile Gln Phe Leu Ala His Leu  
 260 265  
 (2) INFORMATION FOR SEQ ID NO:11:  
 20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1197 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 30 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Kurthia sp.  
 35 (B) STRAIN: 538-KA26  
 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 40 (B) LOCATION: 1..1194  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:  
 45 ATG CAC AGT GAA AAA CAA TTA CCT TGT TGG GAA GAA AAA ATT AAG AAA 48  
 Met His Ser Glu Lys Gln Leu Pro Cys Trp Glu Glu Lys Ile Lys Lys  
 270 275 280  
 GAA CTG GCT TAT TTA GAA GAG ATA TCG CAA AAA CGT GAA CTC GTT TCA 96  
 50 Glu Leu Ala Tyr Leu Glu Glu Ile Ser Gln Lys Arg Glu Leu Val Ser  
 285 290 295  
 ACG GAA TTC GCC GAG CAG CCA TGG CTT ATG ATC AAC GGG TGC AAG ATG 144  
 Thr Glu Phe Ala Glu Gln Pro Trp Leu Met Ile Asn Gly Cys Lys Met  
 300 305 310 315  
 55 CTA AAT CTA GCT TCT AAT AAC TAT TTA GGA TAT GCA GGG GAT GAG CGG 192  
 Leu Asn Leu Ala Ser Asn Asn Tyr Leu Gly Tyr Ala Gly Asp Glu Arg  
 320 325 330

	CTG AAA AAG GCT ATG GTA GAT GCA GTA CAT ACA TAT GGT GCA GGA GCG	240
	Leu Lys Lys Ala Met Val Asp Ala Val His Thr Tyr Gly Ala Gly Ala	
	335 340 345	
5	ACG GCT TCA CGT TTA ATT ATT GGC AAT CAC CCT CTT TAC GAG CAA GCA	288
	Thr Ala Ser Arg Leu Ile Ile Gly Asn His Pro Leu Tyr Glu Gln Ala	
	350 355 360	
10	GAA CAA GCT CTT GTC AAT TGG AAG AAA GCC GAA GCA GGA CTC ATT ATT	336
	Glu Gln Ala Leu Val Asn Trp Lys Lys Ala Glu Ala Gly Leu Ile Ile	
	365 370 375	
15	AAC AGT GGA TAT AAC GCG AAC CTT GGA ATT ATC TCC ACC TTG CTG TCC	384
	Asn Ser Gly Tyr Asn Ala Asn Leu Gly Ile Ile Ser Thr Leu Leu Ser	
	380 385 390 395	
20	CGT AAC GAT ATT ATT TAT AGC GAT AAA TTG AAT CAT GCA AGC ATT GTC	432
	Arg Asn Asp Ile Ile Tyr Ser Asp Lys Leu Asn His Ala Ser Ile Val	
	400 405 410	
	GAT GGA GCT CTC TTA AGC CGT GCA AAG CAT CTA CGC TAT CGT CAT AAT	480
	Asp Gly Ala Leu Leu Ser Arg Ala Lys His Leu Arg Tyr Arg His Asn	
	415 420 425	
25	GAT TTA GAT CAT TTA GAA GCA TTA TTG AAA AAA TCA TCG ATG GAA GCA	528
	Asp Leu Asp His Leu Glu Ala Leu Leu Lys Lys Ser Ser Met Glu Ala	
	430 435 440	
30	CGT AAA TTA ATT GTG ACG GAT ACG GTC TTC AGC ATG GAC GGT GAC TTT	576
	Arg Lys Leu Ile Val Thr Asp Thr Val Phe Ser Met Asp Gly Asp Phe	
	445 450 455	
35	GCT TAT CTT GAA GAC CTT GTT CGG TTA AAA GAA CGT TAT AAC GCT ATG	624
	Ala Tyr Leu Glu Asp Leu Val Arg Leu Lys Glu Arg Tyr Asn Ala Met	
	460 465 470 475	
40	TTA ATG ACA GAT GAA GCA CAC GGA AGC GGC ATC TAC GGT AAA AAC GGT	672
	Leu Met Thr Asp Glu Ala His Gly Ser Gly Ile Tyr Gly Lys Asn Gly	
	480 485 490	
	GAA GGT TAT GCC GGT CAT CTC CAT CTT CAA AAT AAA ATA GAT ATC CAA	720
	Glu Gly Tyr Ala Gly His Leu His Leu Gln Asn Lys Ile Asp Ile Gln	
	495 500 505	
45	ATG GGA ACA TTC AGT AAA GCG CTC GGT TCA TTC GGG GCC TAT GTC GTC	768
	Met Gly Thr Phe Ser Lys Ala Leu Gly Ser Phe Gly Ala Tyr Val Val	
	510 515 520	
50	GGG AAA AAA TGG CTC ATC GAC TAT TTA AAA AAT CGC ATG CGC GGA TTC	816
	Gly Lys Lys Trp Leu Ile Asp Tyr Leu Lys Asn Arg Met Arg Gly Phe	
	525 530 535	
55	ATA TAT TCA ACT GCA CTC CCC CCG GCC ATA CTC GGT GCT ATG AAA ACA	864
	Ile Tyr Ser Thr Ala Leu Pro Pro Ala Ile Leu Gly Ala Met Lys Thr	
	540 545 550 555	

	GCG ATA GAA CTT GTA CAG CAA GAA CCA GAA CGC CGC TCA CTG CTC CAA	912
	Ala Ile Glu Leu Val Gln Gln Glu Pro Glu Arg Arg Ser Leu Leu Gln	
	560 565 570	
5	ACA CAT TCA GAA CAC TTT AGA GAA GAA CTC ACA TAT TAC GGG TTT AAT	960
	Thr His Ser Glu His Phe Arg Glu Glu Leu Thr Tyr Tyr Gly Phe Asn	
	575 580 585	
10	ATT TGT GGA AGT CGA TCA CAA ATT GTT CCT ATC GTC ATC GGG GAA AAC	1008
	Ile Cys Gly Ser Arg Ser Gln Ile Val Pro Ile Val Ile Gly Glu Asn	
	590 595 600	
15	GAA AAA GCG ATG GAA TTT GCC ACA CGT TTG CAG AAA GAA GGA ATT GCA	1056
	Glu Lys Ala Met Glu Phe Ala Thr Arg Leu Gln Lys Glu Gly Ile Ala	
	605 610 615	
20	GCT ATT GCT GTC AGG CCG CCG ACC GTT CCT GAA AAT GAG GCG AGA ATC	1104
	Ala Ile Ala Val Arg Pro Pro Thr Val Pro Glu Asn Glu Ala Arg Ile	
	620 625 630 635	
25	CGT TTT ACT GTA ACA GCT CTC CAC GAT AAA AAA GAT CTT GAT TGG GCA	1152
	Arg Phe Thr Val Thr Ala Leu His Asp Lys Lys Asp Leu Asp Trp Ala	
	640 645 650	
30	GTT GAA AAA GTT TCG ATC ATT GGA AAA GAA ATG GGT GTT ATT	1194
	Val Glu Lys Val Ser Ile Ile Gly Lys Glu Met Gly Val Ile	
	655 660 665	
	TAA	1197

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 398 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met His Ser Glu Lys Gln Leu Pro Cys Trp Glu Glu Lys Ile Lys Lys	
1 5 10 15	
Glu Leu Ala Tyr Leu Glu Glu Ile Ser Gln Lys Arg Glu Leu Val Ser	
20 25 30	
Thr Glu Phe Ala Glu Gln Pro Trp Leu Met Ile Asn Gly Cys Lys Met	
35 40 45	
Leu Asn Leu Ala Ser Asn Asn Tyr Leu Gly Tyr Ala Gly Asp Glu Arg	
50 55 60	
Leu Lys Lys Ala Met Val Asp Ala Val His Thr Tyr Gly Ala Gly Ala	
65 70 75 80	

Thr Ala Ser Arg Leu Ile Ile Gly Asn His Pro Leu Tyr Glu Gln Ala  
 85 90 95  
 5 Glu Gln Ala Leu Val Asn Trp Lys Lys Ala Glu Ala Gly Leu Ile Ile  
 100 105 110  
 Asn Ser Gly Tyr Asn Ala Asn Leu Gly Ile Ile Ser Thr Leu Leu Ser  
 115 120 125  
 10 Arg Asn Asp Ile Ile Tyr Ser Asp Lys Leu Asn His Ala Ser Ile Val  
 130 135 140  
 Asp Gly Ala Leu Leu Ser Arg Ala Lys His Leu Arg Tyr Arg His Asn  
 145 150 155 160  
 15 Asp Leu Asp His Leu Glu Ala Leu Leu Lys Lys Ser Ser Met Glu Ala  
 165 170 175  
 20 Arg Lys Leu Ile Val Thr Asp Thr Val Phe Ser Met Asp Gly Asp Phe  
 180 185 190  
 Ala Tyr Leu Glu Asp Leu Val Arg Leu Lys Glu Arg Tyr Asn Ala Met  
 195 200 205  
 25 Leu Met Thr Asp Glu Ala His Gly Ser Gly Ile Tyr Gly Lys Asn Gly  
 210 215 220  
 Glu Gly Tyr Ala Gly His Leu His Leu Gln Asn Lys Ile Asp Ile Gln  
 225 230 235 240  
 30 Met Gly Thr Phe Ser Lys Ala Leu Gly Ser Phe Gly Ala Tyr Val Val  
 245 250 255  
 35 Gly Lys Lys Trp Leu Ile Asp Tyr Leu Lys Asn Arg Met Arg Gly Phe  
 260 265 270  
 Ile Tyr Ser Thr Ala Leu Pro Pro Ala Ile Leu Gly Ala Met Lys Thr  
 275 280 285  
 40 Ala Ile Glu Leu Val Gln Gln Glu Pro Glu Arg Arg Ser Leu Leu Gln  
 290 295 300  
 Thr His Ser Glu His Phe Arg Glu Glu Leu Thr Tyr Tyr Gly Phe Asn  
 305 310 315 320  
 45 Ile Cys Gly Ser Arg Ser Gln Ile Val Pro Ile Val Ile Gly Glu Asn  
 325 330 335  
 Glu Lys Ala Met Glu Phe Ala Thr Arg Leu Gln Lys Glu Gly Ile Ala  
 340 345 350  
 50 Ala Ile Ala Val Arg Pro Pro Thr Val Pro Glu Asn Glu Ala Arg Ile  
 355 360 365  
 55 Arg Phe Thr Val Thr Ala Leu His Asp Lys Lys Asp Leu Asp Trp Ala  
 370 375 380

Val Glu Lys Val Ser Ile Ile Gly Lys Glu Met Gly Val Ile  
 385 390 395

5 (2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 747 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- 20 (A) ORGANISM: Kurthia sp.  
 (B) STRAIN: 538-KA26

(ix) FEATURE:

- 25 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..744

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

30 ATG AAA CAG CCG AAT TTA GTC ATG CTT CCT GGC TGG GGA ATG GAA AAA 48  
 Met Lys Gln Pro Asn Leu Val Met Leu Pro Gly Trp Gly Met Glu Lys  
 400 405 410

35 GAT GCG TTT CAA CCG TTA ATC AAA CCG CTG TCA GAA GTA TTT CAC CTC 96  
 Asp Ala Phe Gln Pro Leu Ile Lys Pro Leu Ser Glu Val Phe His Leu  
 415 420 425 430

40 TCA TTC ATA GAA TGG AGA GAT ATG AAA ACA CTA AAT GAC TTT GAA GAA 144  
 Ser Phe Ile Glu Trp Arg Asp Met Lys Thr Leu Asn Asp Phe Glu Glu  
 435 440 445

45 CGA GTC ATA GAC ACA ATC GCT TCT ATT GAT GGT CCT GTT TTT TTA CTT 192  
 Arg Val Ile Asp Thr Ile Ala Ser Ile Asp Gly Pro Val Phe Leu Leu  
 450 455 460

50 GGC TGG TCA TTA GGA TCT CTA TTA TCA CTT GAA CTT GTA AGT TCG TAT 240  
 Gly Trp Ser Leu Gly Ser Leu Leu Ser Leu Glu Leu Val Ser Ser Tyr  
 465 470 475

55 CGA GAA AAA ATA AAA GGT TTT ATA CTA ATT GGC GCA ACA AGT CGT TTT 288  
 Arg Glu Lys Ile Lys Gly Phe Ile Leu Ile Gly Ala Thr Ser Arg Phe  
 480 485 490

ACC ACA GGA GAT AAT TAT TCA TTT GGC TGG GAT CCA CGA ATG GTC GAG 336  
 Thr Thr Gly Asp Asn Tyr Ser Phe Gly Trp Asp Pro Arg Met Val Glu  
 495 500 505 510



	CGC ATG AAG AAA CAA CTG CAG CGC AAT AAA GAG AAG ACT TTG ACT TCT	384
	Arg Met Lys Lys Gln Leu Gln Arg Asn Lys Glu Lys Thr Leu Thr Ser	
	515 520 525	
5	TTC TAT GAA GCA ATG TTT TCC GAA GCT GAA AAA GAA GAA GGT TTT TAT	432
	Phe Tyr Glu Ala Met Phe Ser Glu Ala Glu Lys Glu Glu Gly Phe Tyr	
	530 535 540	
10	CAT CAA TTC ATC ACG ACA ATT CAA AGC GAG TTT CAT GGG GAT GAC GTA	480
	His Gln Phe Ile Thr Thr Ile Gln Ser Glu Phe His Gly Asp Asp Val	
	545 550 555	
15	TTT TCG CTT CTT ATA GGT TTG GAT TAT TTA CTT CAG AAA GAT GTT AGA	528
	Phe Ser Leu Leu Ile Gly Leu Asp Tyr Leu Leu Gln Lys Asp Val Arg	
	560 565 570	
20	GTA AAG CTC GAC CAG ATT GAA ACT CCC ATT TTA TTG ATC CAT GGG AGA	576
	Val Lys Leu Asp Gln Ile Glu Thr Pro Ile Leu Leu Ile His Gly Arg	
	575 580 585 590	
	GAA GAC AAA ATT TGT CCA CTC GAA GCC TCA TCT TTC ATT AAA GAA AAT	624
	Glu Asp Lys Ile Cys Pro Leu Glu Ala Ser Ser Phe Ile Lys Glu Asn	
	595 600 605	
25	CTG GGT GGG AAA GCC GAG GTT CAT ATT ATC GAA GGC GCT GGT CAT ATT	672
	Leu Gly Gly Lys Ala Glu Val His Ile Ile Glu Gly Ala Gly His Ile	
	610 615 620	
30	CCA TTT TTC ACA AAA CCA CAG GAA TGT GTG CAG CTT ATA AAA ACA TTT	720
	Pro Phe Phe Thr Lys Pro Gln Glu Cys Val Gln Leu Ile Lys Thr Phe	
	625 630 635	
35	ATT CAA AAG GAG TAC ATT CAT GAT TGA	747
	Ile Gln Lys Glu Tyr Ile His Asp	
	640 645	

(2) INFORMATION FOR SEQ ID NO:14:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

50	Met Lys Gln Pro Asn Leu Val Met Leu Pro Gly Trp Gly Met Glu Lys	
	1 5 10 15	
	Asp Ala Phe Gln Pro Leu Ile Lys Pro Leu Ser Glu Val Phe His Leu	
	20 25 30	
55	Ser Phe Ile Glu Trp Arg Asp Met Lys Thr Leu Asn Asp Phe Glu Glu	
	35 40 45	
	Arg Val Ile Asp Thr Ile Ala Ser Ile Asp Gly Pro Val Phe Leu Leu	

	50		55		60											
	Gly	Trp	Ser	Leu	Gly	Ser	Leu	Leu	Ser	Leu	Glu	Leu	Val	Ser	Ser	Tyr
	65				70					75						80
5	Arg	Glu	Lys	Ile	Lys	Gly	Phe	Ile	Leu	Ile	Gly	Ala	Thr	Ser	Arg	Phe
					85					90					95	
10	Thr	Thr	Gly	Asp	Asn	Tyr	Ser	Phe	Gly	Trp	Asp	Pro	Arg	Met	Val	Glu
			100						105					110		
	Arg	Met	Lys	Lys	Gln	Leu	Gln	Arg	Asn	Lys	Glu	Lys	Thr	Leu	Thr	Ser
			115					120					125			
15	Phe	Tyr	Glu	Ala	Met	Phe	Ser	Glu	Ala	Glu	Lys	Glu	Glu	Gly	Phe	Tyr
	130						135					140				
	His	Gln	Phe	Ile	Thr	Thr	Ile	Gln	Ser	Glu	Phe	His	Gly	Asp	Asp	Val
20	145				150					155						160
	Phe	Ser	Leu	Leu	Ile	Gly	Leu	Asp	Tyr	Leu	Leu	Gln	Lys	Asp	Val	Arg
					165					170					175	
25	Val	Lys	Leu	Asp	Gln	Ile	Glu	Thr	Pro	Ile	Leu	Leu	Ile	His	Gly	Arg
			180						185					190		
	Glu	Asp	Lys	Ile	Cys	Pro	Leu	Glu	Ala	Ser	Ser	Phe	Ile	Lys	Glu	Asn
			195					200					205			
30	Leu	Gly	Gly	Lys	Ala	Glu	Val	His	Ile	Ile	Glu	Gly	Ala	Gly	His	Ile
	210						215					220				
	Pro	Phe	Phe	Thr	Lys	Pro	Gln	Glu	Cys	Val	Gln	Leu	Ile	Lys	Thr	Phe
35	225				230						235					240
	Ile	Gln	Lys	Glu	Tyr	Ile	His	Asp								
					245											

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 831 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Kurthia sp.

(B) STRAIN: 538-KA26

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..828

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

10	ATG ATT GAT AAA CAA TTG TTA AGT AAG CGA TTC AGT GAA CAT GCG AAA Met Ile Asp Lys Gln Leu Leu Ser Lys Arg Phe Ser Glu His Ala Lys	48
	250 255 260	
15	ACA TAT GAT GCA TAT GCC AAT GTT CAA AAA AAC ATG GCG AAA CAA TTA Thr Tyr Asp Ala Tyr Ala Asn Val Gln Lys Asn Met Ala Lys Gln Leu	96
	265 270 275 280	
20	GTG GAT TTG CTC CCT CAA AAA AAC AGC AAA CAA AGA ATT AAC ATC CTT Val Asp Leu Leu Pro Gln Lys Asn Ser Lys Gln Arg Ile Asn Ile Leu	144
	285 290 295	
25	GAA ATT GGC TGC GGT ACT GGT TAC TTA ACC AGG TTA CTC GTT AAT ACA Glu Ile Gly Cys Gly Thr Gly Tyr Leu Thr Arg Leu Leu Val Asn Thr	192
	300 305 310	
30	TTT CCT AAT GCT TCT ATT ACC GCT GTT GAT TTA GCA CCA GGG ATG GTT Phe Pro Asn Ala Ser Ile Thr Ala Val Asp Leu Ala Pro Gly Met Val	240
	315 320 325	
35	GAA GTG GCG AAA GGA ATA ACA ATG GAA GAC CGT GTT ACT TTT TTA TGT Glu Val Ala Lys Gly Ile Thr Met Glu Asp Arg Val Thr Phe Leu Cys	288
	330 335 340	
40	GCT GAT ATC GAA GAA ATG ACG CTT AAT GAA AAT TAC GAC TTA ATT ATT Ala Asp Ile Glu Glu Met Thr Leu Asn Glu Asn Tyr Asp Leu Ile Ile	336
	345 350 355 360	
45	TCT AAT GCA ACG TTT CAA TGG CTG AAT AAT CTT CCT GGA ACC ATT GAA Ser Asn Ala Thr Phe Gln Trp Leu Asn Asn Leu Pro Gly Thr Ile Glu	384
	365 370 375	
50	CAA TTG TTT ACA CGA TTA ACG CCT GAA GGA AAC CTG ATA TTT TCA ACG Gln Leu Phe Thr Arg Leu Thr Pro Glu Gly Asn Leu Ile Phe Ser Thr	432
	380 385 390	
55	TTT GGA ATT AAA ACC TTT CAA GAG CTT CAT ATG TCC TAT GAA CAT GCG Phe Gly Ile Lys Thr Phe Gln Glu Leu His Met Ser Tyr Glu His Ala	480
	395 400 405	
60	AAA GAA AAG CTT CAA CTT TCA ATT GAT AGT TCA CCA GGC CAA CTG TTT Lys Glu Lys Leu Gln Leu Ser Ile Asp Ser Ser Pro Gly Gln Leu Phe	528
	410 415 420	
65	TAC GCT CTA GAA GAA TTA TCC CAA ATT TGT GAA GAA GCA ATC CCT TTT Tyr Ala Leu Glu Glu Leu Ser Gln Ile Cys Glu Glu Ala Ile Pro Phe	576
	425 430 435 440	
70	TCA TCA GCA TTT CCA TTA GAG ATA ACA AAA ATA GAA AAG CTT GAA CTA Ser Ser Ala Phe Pro Leu Glu Ile Thr Lys Ile Glu Lys Leu Glu Leu	624
	445 450 455	

	GAG TAC TTT CAG ACA GTA CGT GAA TTT TTC ACT TCA ATT AAA AAG ATT	672
	Glu Tyr Phe Gln Thr Val Arg Glu Phe Phe Thr Ser Ile Lys Lys Ile	
	460 465 470	
5	GGT GCA GCT AAC AGC AAC AAA GAA AAC TAC TGC CAG CGC CCT TCT TTT	720
	Gly Ala Ala Asn Ser Asn Lys Glu Asn Tyr Cys Gln Arg Pro Ser Phe	
	475 480 485	
10	TTT CGA GAG TTA ATC AAC ATA TAC GAA ACA AAA TAC CAA GAT GAA TCA	768
	Phe Arg Glu Leu Ile Asn Ile Tyr Glu Thr Lys Tyr Gln Asp Glu Ser	
	490 495 500	
15	GGT GTG AAG GCA ACC TAT CAC TGT TTG TTT TTT AAG ATA ATA AAA CAT	816
	Gly Val Lys Ala Thr Tyr His Cys Leu Phe Phe Lys Ile Ile Lys His	
	505 510 515 520	
20	GCC CCC CTA CCC TAA	831
	Ala Pro Leu Pro	

(2) INFORMATION FOR SEQ ID NO:16:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 276 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein

(xi). SEQUENCE DESCRIPTION: SEQ ID NO:16:

35	Met Ile Asp Lys Gln Leu Leu Ser Lys Arg Phe Ser Glu His Ala Lys	1 5 10 15
	Thr Tyr Asp Ala Tyr Ala Asn Val Gln Lys Asn Met Ala Lys Gln Leu	20 25 30
40	Val Asp Leu Leu Pro Gln Lys Asn Ser Lys Gln Arg Ile Asn Ile Leu	35 40 45
45	Glu Ile Gly Cys Gly Thr Gly Tyr Leu Thr Arg Leu Leu Val Asn Thr	50 55 60
	Phe Pro Asn Ala Ser Ile Thr Ala Val Asp Leu Ala Pro Gly Met Val	65 70 75 80
50	Glu Val Ala Lys Gly Ile Thr Met Glu Asp Arg Val Thr Phe Leu Cys	85 90 95
	Ala Asp Ile Glu Glu Met Thr Leu Asn Glu Asn Tyr Asp Leu Ile Ile	100 105 110
55	Ser Asn Ala Thr Phe Gln Trp Leu Asn Asn Leu Pro Gly Thr Ile Glu	115 120 125

	Gln	Leu	Phe	Thr	Arg	Leu	Thr	Pro	Glu	Gly	Asn	Leu	Ile	Phe	Ser	Thr	
	130						135					140					
5	Phe	Gly	Ile	Lys	Thr	Phe	Gln	Glu	Leu	His	Met	Ser	Tyr	Glu	His	Ala	
	145					150					155					160	
	Lys	Glu	Lys	Leu	Gln	Leu	Ser	Ile	Asp	Ser	Ser	Pro	Gly	Gln	Leu	Phe	
				165						170					175		
10	Tyr	Ala	Leu	Glu	Glu	Leu	Ser	Gln	Ile	Cys	Glu	Glu	Ala	Ile	Pro	Phe	
			180						185					190			
	Ser	Ser	Ala	Phe	Pro	Leu	Glu	Ile	Thr	Lys	Ile	Glu	Lys	Leu	Glu	Leu	
15			195					200					205				
	Glu	Tyr	Phe	Gln	Thr	Val	Arg	Glu	Phe	Phe	Thr	Ser	Ile	Lys	Lys	Ile	
	210						215					220					
20	Gly	Ala	Ala	Asn	Ser	Asn	Lys	Glu	Asn	Tyr	Cys	Gln	Arg	Pro	Ser	Phe	
	225					230					235					240	
	Phe	Arg	Glu	Leu	Ile	Asn	Ile	Tyr	Glu	Thr	Lys	Tyr	Gln	Asp	Glu	Ser	
				245						250					255		
25	Gly	Val	Lys	Ala	Thr	Tyr	His	Cys	Leu	Phe	Phe	Lys	Ile	Ile	Lys	His	
			260						265					270			
	Ala	Pro	Leu	Pro													
30			275														